

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	3859	83.5	879	12	Q9WU9	Q9wiu9 human immun
2	3670	79.4	876	12	Q9YKB2	Q9Ykb2 human immun
3	3664	79.3	872	12	O57072	O57072 human immun
4	3664	79.3	871	12	O57073	O57073 human immun
5	3654	79.1	900	12	Q9GN28	Q9gnz8 human immun
6	3650	79.0	871	12	O57074	O57074 human immun
7	3632	78.6	863	12	O77377	O77377 human immun
8	3577	77.4	860	12	Q93091	Q93091 human immun
9	3414	73.9	877	12	O76576	O76576 human immun
10	3381	73.2	876	12	O79670	O79670 human immun
11	3283	51.6	840	12	Q9GNX7	Q9gnx7 human immun
12	2370	51.3	863	12	O40360	O40360 human immun
13	2356	51.0	860	12	Q9QRW2	Q9qrw2 human immun
14	2356	51.0	862	12	Q9WLG8	Q9wlg8 human immun
15	2342	50.7	859	12	O9WLJ1	O9wlj1 human immun
16	2338	50.6	537	12	Q9IED8	Q9ied8 human immun
17	2334	50.5	861	12	O73366	O73366 human immun
18	2334	50.5	859	12	O76128	O76128 human immun
19	2332	50.5	864	12	Q70210	Q70210 human immun

CHAIN 132
CHAIN 164

POTENTIAL.
GP120.

FT CHAIN 302 335 V3.
 FT CHAIN 395 427 V4.
 FT CHAIN 458 475 V5.
 FT CHAIN 525 879 GP41.
 SQ SEQUENCE 879 AA: 99099 MM: BC6422ADDF1A3409 CRC64;

Query Match 83.5%; Score 3859; DB 12; Length 879;
 Best Local Similarity 82.5%; Pred. No. 0;
 Matches 730; Conservative 58; Mismatches 79; Indels 18; Gaps 8;

QY 1 MYTMRAMGRNRKRLGILYVALIIPCLSSSOLATYAGVWEDAPVLEFCSADNL 60
 DB 1 MYTMRAMGRNRKRLGILYVALIIPCLSSSOLATYAGVWEDAPVLEFCSADNL 60
 QY 61 TSEKHNWASQACVPTDPTPEHYLLNTDNTDNFNTENTVEMODIISLMDOSLPCV 120
 DB 61 TSEKHNWASQACVPTDPTPEHYLLNTDNTDNFNTENTVEMODIISLMDOSLPCV 120
 QY 121 QMTFICIONCTDIKNNNTSGTENRST-----SEENMKCEPNTTVLADKREKCA 173
 DB 121 QMTFICIONCTDIKNNNTSGTENRST-----SEENMKCEPNTTVLADKREKCA 180
 QY 174 LEYVSDLEFLADNNTNTNTLLINCSNTTIKQACPVSEPIPIYICAPAGAFKCSNA 233
 DB 174 LEYVSDLVK-TDN---STWYTLINCSNTTIKQACPVSEPIPIHYCAPAGAFKRCNT 236
 QY 234 EENGKCGKNSIVYCTHRIKFTVSTQILNGLTSEKIRINGKNISDGKNIIVTLSSD 293
 DB 237 EENGKCGKNSIVYCTHRIKFTVSTQILNGLTSEKIRINGKNISDGKNIIVTLSS 296
 QY 294 IEITCVPRGNNOYEMKIGPMATSMALGTSNRSVAYCOQNTTEWEKAKNAEAYL 353
 DB 297 ISMTCEPRGNH-TVOEMKIGPMATSMALGTSNRSVAYCOQNTTEWEKAKNAEAYL 355
 QY 354 ELINTEGNT-TMIPNRSODGSDVEYTHLFHNGHEFFYCANTSENFYTLONGTNCNT 412
 DB 356 ELVNTSNKTYVMIFNOSNDG-DPEVTHLFHNGHEFFYCANTSENFYTLONGTNCNT 414
 QY 413 OSINANG--MIPCKLKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNT 470
 DB 415 ONNNTAHNDTQIPCKLKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNT 474
 QY 471 NTFPRIGDKMDIWFENELFKYVAVKPSVAPPIAPVIGTGRHRRKRAVGKMLFG 530
 DB 475 TTFPRVGDMDIWFTELFKRYVAVKPSVAPPIAPVIGTGRHRRKRAVGKMLFG 534
 QY 531 VLSAAGSTGAATATLVQVTHSVKGIYOQDNLRAIOAOQELRLSYWGIROLBARL 590
 DB 535 VLSAAGSTGAATATLVQVTHSVKGIYOQDNLRAIOAOQELRLSYWGIROLBARL 594
 QY 591 ALETLIOQOOLNMGCGRLICTSVKMETWNTNTIN--QIWGNTTWOEQOQIDNV 648
 DB 595 ALETLIOQOOLNMGCGRLICTSVKMETWNTNTIN--QIWGNTTWOEQOQIDNV 654
 QY 649 SSTIYEEOKAOVOEOEKEKLEDEWASLMMWLDITKMYKIMIIYVALIGRIYV 708
 DB 655 SSTIYEEOKAOVOEOEKEKLEDEWASLMMWLDITKMYKIMIIYVALIGRIYV 714
 QY 709 MIVLNVNIRGQVPLSLQIPTRQOSEAETPGRTGEGGDEGRPLIPPOGELPLLYT 768
 DB 715 MIVLNVNIRGQVPLSLQIPTRQOSEAETPGRTGEGGDEGRPLIPPOGELPLLYT 774
 QY 769 DLRTIILMSYHLSTLIGTQVISHRLGIMIGOKIIDACRICAAVHYWLOELOKSA 828
 DB 775 DLRTIILMSYHLSTLIGTQVISHRLGIMIGOKIIDACRICAAVHYWLOELOKSA 834
 QY 829 TSLIDTFAVAVANTDDIIGIOLRGRIINPRVRGPRSTL 873
 DB 835 TSLIDTFAVAVANTDDIIGIOLRGRIINPRVRGPRSTL 879

RESULT 2

QYK92 ID QYK92 PRELIMINARY; PRT: 876 AA.
 AC QYK92
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE GP160 PROCURSOR.
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroviruses; Retroviridae; Lentiviruses.
 OX NCBI_Taxid=11676;
 RN [1]
 RP SEQUENCE FROM A.A.
 RC STRAIN=CA9;
 RX MEDLINE=9923950; PubMed=10207543;
 RA Janssens W., Heyndrickx L., Van der Auwerda G., Nkengasong J.,
 RA Belmans E., Verheijden K., Copens S., Williams B., Fransen K.,
 RA Peeters M., Ndumba P., Delaporte E., Van der Groen G.,
 RT "Interpatient genetic variability of HIV-1 group O.";
 RL AIDS 13:41-48 (1999).
 DR EMBL: X96522; CA65306.1;
 DR INTERPRO: IPR000328;
 DR INTERPRO: IPR00777;
 DR PRAM: PR00516; GR120; 1.
 DR PRAM: PR00517; GP41; 1.
 KW signal.
 FT SIGNAL 1 29
 FT CHAIN 30 524
 FT CHAIN 132 156
 FT CHAIN 158 199
 FT CHAIN 301 335
 FT CHAIN 394 425
 FT CHAIN 458 474
 FT CHAIN 525 876
 FT CHAIN GP41.
 SQ SEQUENCE 876 AA: 98879 MM: B084D1B2EC656C9F CRC64;

Query Match 79.4%; Score 3670.5; DB 12; Length 876;
 Best Local Similarity 80.1%; Pred. No. 1.5e-295;
 Matches 707; Conservative 63; Mismatches 96; Indels 17; Gaps 12;

QY 1 MYTMRAMGRNRKRLGILYVALIIPCLSSSOLATYAGVWEDAPVLEFCSADNL 60
 DB 1 MYTMRAMGRNRKRLGILYVALIIPCLSSSOLATYAGVWEDAPVLEFCSADNL 60
 QY 61 TSEKHNWASQACVPTDPTPEHYLLNTDNTDNFNTENTVEMODIISLMDOSLPCV 120
 DB 61 TSEKHNWASQACVPTDPTPEHYLLNTDNTDNFNTENTVEMODIISLMDOSLPCV 120
 QY 121 QMTFICIONCTDIKNNNTSGTENRST-----SEENMKCEPNTTVLADKREKCA 179
 DB 121 QMTFICIONCTDIKNNNTSGTENRST-----SEENMKCEPNTTVLADKREKCA 180
 QY 180 LFTLANNTNT--TWYTLINCSNTTIKQACPVSEPIPIYICAPAGAFKCSAEPN 236
 DB 181 LFTLANNTNT--TWYTLINCSNTTIKQACPVSEPIPIYICAPAGAFKCSAEPN 238
 QY 237 GTCKGNSISVYCTHRIKFTVSTQILNGLTSEKIRINGKNISDGKNIIVTLSSDI 296
 DB 239 GTCKGNSISVYCTHRIKFTVSTQILNGLTSEKIRINGKNISDGKNIIVTLSSDI 298
 QY 297 ACVRPGRNNOYEMKIGPMATSMALGTS-NRSVAYCOQNTTEWEKAKNAEAYL 355
 DB 299 ACVRPGRNNOYEMKIGPMATSMALGTS-NRSVAYCOQNTTEWEKAKNAEAYL 357
 QY 356 INTBENTMTMIPNRSODGSDVEYTHLFHNGHEFFYCANTSENFYTLONGTNCNT 415
 DB 358 INTBENTMTMIPNRSODGSDVEYTHLFHNGHEFFYCANTSENFYTLONGTNCNT 415
 QY 416 NSANG--MIPCKLKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNT 472
 DB 416 TNDNGTMDIPCKLKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNT 475

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: January 25, 2001, 10:51:25 ; Search time 15.7 Seconds
(without alignments)
1795.716 Million cell updates/sec
Title: US-08-911-824-61
Perfect score: 4622
Sequence: 1 MIVTWRMCKRNRKILYI.....GRGILNPRVRQGFERSLL 873
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 88757 seqs, 32294092 residues
Total number of hits satisfying chosen parameters: 88757
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SWISSPROT_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2318.5	50.2	854	1 ENV_SIVCZ	P17281 chimpanzee
2	2290	49.5	856	1 ENV_HV12H	P05881 human immun
3	2287	49.5	853	1 ENV_HV12H	P04581 human immun
4	2264.5	49.0	859	1 ENV_HV12H	P04583 human immun
5	2263.5	49.0	855	1 ENV_HV12H	P04580 human immun
6	2261	48.9	863	1 ENV_HV12H	P05882 human immun
7	2259	48.9	847	1 ENV_HV12H	P19550 human immun
8	2247.5	48.6	868	1 ENV_HV12H	P05879 human immun
9	2247	48.6	843	1 ENV_HV12H	P35961 human immun
10	2242	48.5	853	1 ENV_HV12H	P12487 human immun
11	2239.5	48.5	852	1 ENV_HV12H	P19549 human immun
12	2238.5	48.4	848	1 ENV_HV12H	P20871 human immun
13	2233.5	48.3	861	1 ENV_HV12H	P31819 human immun
14	2231.5	48.3	856	1 ENV_HV12H	P31872 human immun
15	2228.5	48.2	852	1 ENV_HV12H	P12488 human immun
16	2214.5	47.9	847	1 ENV_HV12H	P05880 human immun
17	2212.5	47.9	846	1 ENV_HV12H	P18799 human immun
18	2211.5	47.8	856	1 ENV_HV12H	P03375 human immun
19	2211.5	47.8	856	1 ENV_HV12H	P05877 human immun
20	2209.5	47.8	856	1 ENV_HV12H	P03376 human immun
21	2209	47.8	851	1 ENV_HV12H	P04582 human immun
22	2208.5	47.8	856	1 ENV_HV12H	P04578 human immun
23	2207	47.7	867	1 ENV_HV12H	P12489 human immun
24	2205.5	47.7	856	1 ENV_HV12H	P070626 human immun
25	2205.5	47.7	861	1 ENV_HV12H	P03377 human immun
26	2197	47.5	855	1 ENV_HV12H	P20888 human immun
27	2194.5	47.5	856	1 ENV_HV12H	P05878 human immun
28	2192.5	47.4	856	1 ENV_HV12H	P04624 human immun
29	2190	47.4	865	1 ENV_HV12H	P04579 human immun
30	2189	47.4	855	1 ENV_HV12H	P03378 human immun
31	2180.5	47.2	853	1 ENV_HV12H	P19551 human immun
32	1274	27.6	885	1 ENV_SIV54	P12452 simian immu
33	1269.5	27.5	856	1 ENV_HV2NZ	P05883 human immun

RESULT 1									
ENV_SIVCZ	ENV_SIVCZ	STANDARD;	PRT;	854 AA.					
AC	P17281;								
DT	01-AUG-1990 (Rel. 15, Created)								
DT	01-AUG-1990 (Rel. 15, Last sequence update)								
DT	15-JUL-1999 (Rel. 38, Last annotation update)								
DE	ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE								
DE	GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].								
GN	ENV.								
OS	Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).								
OC	Viruses; Retroviral viruses; Retroviridae; Lentivirus.								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE-90259077; PubMed-2188136;								
RT	Huet T., Chevillon R., Meyerhans A., Roelants G., Wain-Hobson S.;								
RL	"Genetic organization of a chimpanzee lentivirus related to HIV-1.";								
CC	Nature 345:356-359(1990)								
CC	- SIMILARITY: STRONG, WITH HIV-1 ENV POLYPROTEIN.								
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on its								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/								
CC	or send an email to license@isb-sib.ch).								
CC	-----								
CC	EMBL: X52154; CAA36407.1; --								
DR	PIR: S09990; VCL:SL.								
DR	HIV: X52154; ENVSCBP2.								
DR	INTERPRO: IPR000328; --								
DR	INTERPRO: IPR000777; --								
DR	PFAM: PF00516; GP120; 1.								
DR	PFAM: PF00517; GP41; 1.								
KW	AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;								
KW	Signal.								
FT	SIGNAL	1	30						
FT	CHAIN	31	500						
FT	CHAIN	501	854						
FT	TRANSMEM	501	517						
FT	TRANSMEM	675	693						
FT	TRANSMEM	805	821						
FT	CARBOHYD	134	134						
FT	CARBOHYD	140	140						
FT	CARBOHYD	143	143						
FT	CARBOHYD	154	154						
FT	CARBOHYD	158	158						
FT	CARBOHYD	186	186						
FT	CARBOHYD	195	195						
FT	CARBOHYD	239	239						
FT	CARBOHYD	260	260						
FT	CARBOHYD	267	267						
FT	CARBOHYD	274	274						
FT	CARBOHYD	299	299						
FT	CARBOHYD	331	331						

pending ~~AT~~ removed

P19503 simian immu
Q02837 simian immu
P24105 human immu
P27977 simian immu
P20872 human immu
P04577 human immu
P05885 simian immu
P18094 human immu
P18040 human immu
P05886 human immu
P17755 human immu
P27757 simian immu

FT	CARBOHYD	336	336	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	351	351	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	356	356	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	384	384	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	392	392	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	426	426	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	432	432	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	446	446	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	450	450	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	601	601	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	608	608	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	628	628	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	854 AA;	95803 MW;	2E249AFAD4F29B3 CRC64;	
Query Match					
50.2%; Score 2318.5; DB 1; Length 854;					
Best Local Similarity 50.9%; Pred. No. 1.1e-160;					
Matches 452; Conservativity 148; Mismatches 235; Indels 53; Gaps 16;					
QY	5	MRAMGKRR---	KIGILYIV-	MALIIPLSSQLYATVYAGVPWEDAAAPVLFASDANL	60
DB	1	MKYMKKKRDN	SLSIITII	ILLPCL-TSELMVTVYGVVPWHDADPVLFASDAKA	59
QY	61	TSTKHNWASQ	ACVPTDPT	PHEYLTLNVTDNENINWYEQMEDIIISLWQSLKPCI	120
DB	60	HSTEAHNINAT	ACVPTDPS	QFVLPNVIESFNMMKNMVDQMHEDIISLWQSLKPCV	119
QY	121	QMTFMCQNMCT	DKNNNTSG	TNRSSSENPMKTCFEFNITTVLKOKKKQALFYVSDL	180
DB	120	KLPLCVTLQCS	KANFSQAN	LNTQSSPPLEMKNSFNVTIELRDKKQVYSLFYVEDV	179
QY	181	TKLADNNNT	MTYTLNCS	NTTIKQCPKVSFPIPIYCAPAGYAFKCNSEFNGTGR	240
DB	180	VNLGNENT---	YRLNCNTAI	TQACPKTSFPIPIHYCAPAGFALLKNDKDFSGKG	236
QY	241	CSNISVVTCH	GIKPTVST	QIILNGTSLKKEIRKIMKNISDSKNIIVTLSSDIETCVR	300
DB	237	CTNVSTVHC	HQKPVVTV	TQLLINGSIAEGNITVRVENSKNTDVMVLQVVEAVSLNCR	296
QY	301	PGNNTVQEM	KICP-NAWY	SMALGTSGNRSRVACQYNTTEWELKNTAERYLELINT	359
DB	297	PGNN-TGEV	QIGPMTFY	NIENVGDTSR--AYCKINGTWTNRTVEEVKALATSSNRT	353
QY	360	EGNTTIFNR	SQDSVDEV	THLFHENGGEFFYCYNTSEMFNYTFLCNGTNCNNTQSIANSAN	419
DB	354	AAITL--NRA	-SGGDPV	THHMCNGGEFFYCNYSQIF-----TDNITNGI	397
QY	420	GMTPKLKQV	RSWMRGSG	GLYAPPIGNLTCISHTGMLQMDAP-WNKTEH-TFRPIG	477
DB	398	IILPCIRQI	VSSWMVRG	IRGYIAPPINGNITCSNITGLLTSDTPVTNNSGNLTFRPTG	457
QY	478	GMKMDIWRN	ELFKYKVR	KVPFSAVPTPIARPVIGTGTREKRAV-GLGLMFLGVLISAAG	536
DB	458	GNMKOLWRS	ELYKVKVRI	EPLSVAPTAKRHRTVAKQDKRAAFGLGALFGLFGAAG	517
QY	537	STMGAAAT	ALTATVTH	SIVIKIGVQQQNLRAIQAEQLRLSVWGIQRLARLALLETLI	596
DB	518	STWGAATV	LTVAQRLS	GISVQQQNLKAEIAEQHLLQLSIWGVKQLQARLAVERYL	577
QY	597	QNOQLNLWG	CKGRLCYTS	VKNWETWNTNINQINGNLTWQEWDOQIDNVSSSTIYEI	656
DB	578	QDOQILGLW	CGSKAVCYT	TPVNNNSWPGSNSTDIDWNLTWQWDKLVSNYTKIFGLL	637
QY	657	OKAQVOEQE	NEKKLEDE	WASLWNLDTKWLWYIKIAIIVGALIGVRIYVIMLVNLR	716
DB	638	EEAQSQEK	NERDELLE	QWASLWNLDTKWLWYIKIFLWAVGGIIGRILMTVFSVVR	697
QY	717	NIRQGYPLS	QLOIPTRQ	QSAETPGRTGEGGDEGRPLRPSQGFPLPYLTLRLTIILW	776
DB	698	RVQGYSPLS	QLOIPLVQ	REQRGLGEIDEGGDEGRSRLVVEGCLPLWDDLRLGIW	757
QY	777	SYHLLNSL	ISGTQTVI	SHLRGLW-----ILGQKIIDACR-----ICAAVHYWYQELQ	825

DB	758	SYOGLTS-----	LACNVWRQLTKTGLHLSRLRRLRCLLGGIIQYWGKELK	806
QY	826	KSATSLIDTFA	VAVANWTDI	ILIGLQRLGRGILNIPRRVROGPERILL 873
DB	807	ISALSUDATA	IAVAEGTDRI	IEAFQVTLRIIRNIPRIROGLERALL 854
RESULT 2				
ENV_HV12H	STANDARD;	PRT;	856 AA.	
AC	P05881;			
DT	01-NOV-1988	(Rel. 09, Created)		
DT	01-NOV-1988	(Rel. 09, Last sequence update)		
DT	15-JUL-1999	(Rel. 38, Last annotation update)		
DE	ENVELOPE POLYPROTEIN GP160	PRECURSOR [CONTAINS: EXTERIOR MEMBRANE		
DE	GLYCOPROTEIN (GP120);	TRANSMEMBRANE GLYCOPROTEIN (GP41)].		
GN	ENV.			
OS	Human Immunodeficiency virus type 1 (Zaire H321 isolate) (HIV-1).			
OC	Viruses; Retroviral viruses; Retroviridae; Lentivirus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-89228766; PubMed-2713163;			
RA	Srinivasan A., York D., Butler D. Jr., Jannoun-Nasr R., Getchell J.,			
RA	McCormick J., Ou C.Y., Myers G., Smith T., Chen E.;			
RT	"Molecular characterization of HIV-1 isolated from a serum collected			
RT	in 1976; nucleotide sequence comparison to recent isolates and			
RT	generation of hybrid HIV.;"			
RL	AIDS Res. Hum. Retroviruses 5:121-129(1989).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M15896; AAB53948.1; .			
DR	PIR; A44963; A44963.			
DR	HIV; M15896; ENV32321.			
DR	INTERPRO; IPR000328; .			
DR	INTERPRO; IPR000777; .			
DR	PFAM; PF00516; GP120; 1.			
DR	PFAM; PF00517; GP41; 1.			
KW	AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;			
KW	Signal.			
FT	SIGNAL	1	29	
FT	CHAIN	30	511	EXTERIOR MEMBRANE GLYCOPROTEIN.
FT	CHAIN	512	856	TRANSMEMBRANE GLYCOPROTEIN.
FT	DISULFID	53	73	BY SIMILARITY.
FT	DISULFID	118	206	BY SIMILARITY.
FT	DISULFID	125	197	BY SIMILARITY.
FT	DISULFID	130	153	BY SIMILARITY.
FT	DISULFID	219	248	BY SIMILARITY.
FT	DISULFID	229	240	BY SIMILARITY.
FT	DISULFID	297	331	BY SIMILARITY.
FT	DISULFID	379	445	BY SIMILARITY.
FT	DISULFID	386	418	BY SIMILARITY.
FT	CARBOHYD	87	132	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	132	132	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	152	152	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	156	156	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	183	183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	198	198	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	242	242	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	263	263	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	277	277	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	302	302	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	332	332	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	339	339	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	384	384	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	390	390	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	394	394	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	400	400	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	403	403	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	406	406	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	411	411	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	445	445	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	458	458	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	459	459	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	462	462	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	608	608	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	613	613	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	622	622	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	634	634	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	SEQUENCE	853 AA:	96721 MR:	F9CD864DAAD0D07A5	CRG64;

Query Match 49.5%; Score 2287; DB 1; Length 853;

Best Local Similarity 51.1%; Pred. No. 2e-158;

Matches 450; Conservative 157; Mismatches 234; Indels 40; Gaps

QY	5	MRWAG-KRN-----RKIGILYIVMALIIPCUSSSQLATYATYAGVPVWEDAAPVLFCASDAN	59
DB	1	MRAGIERNCQNKKWGIN--LLGILMTCASADNLATVTVYGVVVKETATTLFCASDASD	58
QY	60	LTSTKHNVWASOACVPTDPTPHEYLTLTNVDFNFWENWVQOMQEDIIISLWDSQSLKPC	119
DB	59	SYTEAHNIWATACVPTDPNPOEIALENTFENMWKNWVQOMHEDIISLWDSQSLKPC	118
QY	120	IOMTFMCIQNCYD-IKNNNTSGTENRISSEPMKTCFENITVTLVKDKKKQALFYVS	178
DB	119	VKLTPCLVTLNCSDLRNNGTGM--NNVITEEGMNCSEFNTVTLKDKKQVYALFYRL	176
QY	179	DLTKLADNNTN-TMYTLINCNTTIQCAPKVSFPPIPIYICAPAGYAFKCNSEAFNG	237
DB	177	DIVIPDNDSSTNTYRLINCNTSAITQCAPKVSFPPIPIHYCAPAGFAILKCRDKKFG	236
QY	238	TGCSNLSVVTCTHGIKPTVSTOLLNGTSLKSKIRIMGNISDSKNIIVTLSSDIET	297
DB	237	TGPTNTSVTQCTHGIKPPVSTQLLNGSLAEVEIIRSENUTNNNAKNIHAHLSVESVKT	296
QY	298	CVRPGNNQTVQEMKIGPMWYSMALGTGNSRSRV--AYCOYNTTEWEKALKNTAEYRLEL	355
DB	297	CARYQN-TRQTPIG----LGOSLYTRSRSIIQGAHCNISRAQSKTLQQVARKLGLT	351
QY	356	INNTGNTMIFNRSDGSDVETHLFNCHGEFFCYNTSEMPNYTFYFLNGNCNNNTQSI	415
DB	352	LNR-----TIIKFPSSGGDEPETHSFNCGGGEFFCYNTSGLFNSTWNISAWN-NITES	405
QY	416	NSANGMI--PCKLKQVYRSMWGGSLYAPPIGNLTCTSHITGMILQWDAPNKNTE-N	472
DB	406	NSTNTNITLQCRKIQIK-MVAGRKAIYAPPIERNLTCSNITGLLTLDGGINNSTNET	464
QY	473	FRPIGGDKDIWENELFKYKVVVKPFSVAPTIAPRPVIGTGHREKRAVGLGMFLGLVL	532
DB	465	FRGGGDMRNWSELYKXKVVQIEPLGVAPTRAKRRV-----ERKRAIGLGMFLGL	520
QY	533	SAAGSTMGAATAATLVQTHSVIKGIVQOQDNLRAIQAOQLRLLSVWGIQRILARLLAL	592
DB	521	GAAGSTMGARSVTLTQVQARQLMSGIVQOQDNLRAIEAQHLLQTLVWGIKLOARILAV	580
QY	593	ETLIQNOQLNLNGCKGRLLICYTSVKWNETWRNTNINQWHLTWQEDQOQIDNVSSFI	652
DB	581	ERYLKQDQLLGIWGGSKHCTCTTVPNWSWSN-RSLNTEWQMTWMEEREIDNTGLI	639
QY	653	YEEIQAAQVOEQNEKRLLEDWASLNNWLDITTKWLTKIAIIVGALIGRIVMAYL	712
DB	640	YSLIERSQTOEKNKEKLELDKWSLNNWFSITQWLWIKIFIMIIIGGLIGLRIVFAYL	699
QY	713	NLYRNTIRSQYQLSLQIPTRTQQSEATPGTGGGGDEGRPLIPISQGFPLLYLTDLRT	772

[illegible]

QY	664	BQNEKKLEDEWASLWNNWLDITTKWLTWIKIAIIIVGALIGVRIWVLNVLNVRNIQGYQ	720
Db	653	EKNEQELLELDKWSLWNNWENITQWLWIKIFIMIVGGLGLGRIVFAVLSLVNVRVQGY	712
QY	724	PLSLQIPTRQSQEAETPGTGGGDDGGRPLIPSPGPFUPLYTLRLTIILWSVHLLSN	783
Db	713	PLSFQILLPAPRPDRPGIEEGEGGERDSIRLVNGFSALIWDDLNLCLFSYHRLRD	772
QY	784	LISGTQTVSHIRLGLWILWLGKIIDACRICAATVHYWLOEQKSAVSLDITFAVAVANT	843
Db	773	LILIAARIVE-----LLGRGWEALKYLNWLLQYWSRELRNSASLLDTTIAIAVAGT	825
QY	844	DDLILGIORLGRGIILNIPRVVQGFERSLL	873
Db	826	DRVIEIVRTYRAVLNVPTRIQGLERLLL	855

RESULT 6
 ENV_HV128
 ID ENV_HV128 STANDARD; PRT; 863 AA.
 AC P05982;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 15-JUL-1998 (Rel. 09, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENVELOPE POLYPEPTIDE GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
 GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].
 GN ENV.
 OS Human immunodeficiency virus type 1 (Z-84 isolate) (HIV-1).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88281278; PubMed=3395517;
 RA Yorno J., Josephs S.F., Reitz M.S. Jr., Zagury D., Wong-Staal F.,
 RA Gallo R.C.;
 RT "Nucleotide sequence analysis of the env gene of a new Zairian
 RT isolate of HIV-1";
 RL AIDS Res. Hum. Retroviruses 4:165-173(1988).
 CC -!- MISCELLANEOUS: THE Z-84 ISOLATE WAS TAKEN FROM A 54 YEAR-OLD
 CC ZAIRIAN MALE.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on
 CC use by non-profit institutions as long as its content is in no
 CC modified and this statement is not removed. Usage by and for commerc
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announc>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J03653; AAA44684.1; -
 CC HIV: J03653; ENV5JY1.
 CC INTERPRO: IPR000328; -
 CC INTERPRO: IPR000777; -
 CC PFAM: PF00516; GP120; 1.
 CC PFAM: PF00517; GP41; 1.
 CC AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
 CC Signal.
 CC

FT	1	29
FT	30	518
FT	519	863
FT	DISULFID	53
FT	DISULFID	118
FT	DISULFID	125
FT	DISULFID	130
FT	DISULFID	230
FT	DISULFID	240
FT	DISULFID	308
FT	DISULFID	388
FT	DISULFID	395
FT	CARBOHYD	87
FT	CARBOHYD	129
FT	CARBOHYD	136
FT	CARBOHYD	142
FT	CARBOHYD	143

Query Match 48.98; Score 2261; DB 1; Length 863;
 Best Local Similarity 49.5%; Pred. No. 1.6e-156;
 Matches 442; Conservative 152; Mismatches 245; Indels 54; Gaps 15;

QY 5 MRAMGKRRN-----KLGILYVMAIIICPLSSQIATYAGVPWEDAAAPVLCASDAN 59
 DB 1 MRVAGIRMYOHLWKGLN--LGLILMTCVSAEDLWTVYGVVWKEATTLTFCASDAK 58

QY 60 LTSTEKNHNVASQCVPTDPTTPEHYLLTYNDNFNINWENTYVEQMEDIIISLWQSLKPC 119
 DB 59 SYEPEAHNIWATHACVPTDPTPREIEMENVTENFMKNMNVQEMHEDIIISLWQSLKPC 118

QY 120 IQMTFMCIQMCTDKNNNTSGTNRSSSE-----NPMKTCENITTVLKDKKREKQAL 174
 DB 119 VKLTPCLVTLNAGNKTNGNTTNGEQEMKEMKCSFNITTVISDKKKQVHAL 178

QY 175 FYVSDLAKLADNNTNTM-----YTLNCSTTIKQACPKVSEPIPIYCAAGVAIF 228
 DB 179 FYRLOVFPIDDDNSANTNTWYRLNLCNTSAITQACPKVTEPIPIHYCAPAGFAIL 238

QY 229 KCNSAEFNGTKCNSISVVTCTHGIKPTVSTQLIINGTLSEKIRIMKNSDSGKNIIY 288
 DB 239 KCDDKKGNGTGPCKVSIVQCTHGRPVSTQLLINGSIAEEIIIRSENITNNVKIIV 298

QY 289 TLSSDIEITCVRPNQNTVQEMKIGPMWYSMALGTGNSRSRV-----AYCQYNTPEWE 342
 DB 299 HLNESVEINCRPNKTRQSTPIG-----LQALYITTRIKGDIRQAYCNISAAWN 350

QY 343 KALANTAERYLELNNTEGNTMIFNRSQDSQDVETHLHFNCHGEFFYCNTSEMYTFP 402
 DB 351 KTLQOVAKKLDGLLN----QTTIIF-KPPAGGDPEITTHSFNCGGEFFYCNTSLFNSTW 405

QY 403 LCNCTNCTNQSINSANGM-IPCKLKQVVRWSRGGSGLYAPPPIPGNLTCTSHITGMILQ 461
 DB 406 --NSTWN--DTLNSGKTIKPCRIKQIINNMGVGRVAPPIEGLKCTNSITGLLT 462

QY 462 MDAPWNKTEN--TFRPIGDMKDIWRNELKYKYVVRVFPFVAPTPPIAPVIGTGTREKR 520
 DB 463 RDGVNNSSTNETFRPGGDMKDNRNELYKYKYVRIEPIPLGIAPTRAKRRVY-----EREK 518

QY 521 AVGLGMLFLGLVLSAAGSTMGAAATLVQHSVILKIGVQQDNLRLAIQAQOELLRLSVW 580
 DB 519 AIGLGAVFLGLGAAGSTMGAVSALTQCARQLLSGIVQQQNNLLRAEIAEQHMLQTVW 578

QY 581 GIORLARLLAELTILQIQOQLLNMGCKRLICYTSVKWNETWNTNINQINLWQOE 640
 DB 579 GIKLOARVLAVESYLKQDQLLGIWCGSGRICTTTVPWNSWSN-KSLEIWNNTWIE 637

QY 641 WDOQIDNVSTIYEIEIOKAQVOQSENEKKLELDEWASLNNWLDITKWLWYIKITAILV 700
 DB 638 WEREIDNTGVISLIENSQIQEKNQDQLLQDKWASLWNWFSITKWLWIKIFIMIVG 697

QY 701 ALIGVRIVMIVNLVRNIQOYQPLSLQIPTRQOSEAETPGRTGEGGDEGRPLIPSPQ 760
 DB 698 GLIGLRIVFVLSLVNRVQSGYSPSLFTLLPAPRDPGPGIEEGEGEQGRSIRLVN 757

QY 761 GFPLPLTDTURTIIILSYHLSNLISGTQVISHRLGLWILGOKIIDACRICAAVHYW 820
 DB 758 GFSALFWDLLRNLCLFSYHRLDLILIAITRIVE-----LLGRGWFAIKYLSLQYW 810

QY 821 LOELQKATSIDTFVAVANWTDIILGLORLGRGILNIPRRVROGFERSLL 873
 DB 811 TQELKNSFISLNNATAVAEGTDRIELIRRAFRVLHPIRRVROGLERALL 863

RESULT 7
 ENV_HV1S1
 ID ENV_HV1S1 STANDARD; PRT; 847 AA.
 AC P19550;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
 DE GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)]
 GN ENV.
 OC Human immunodeficiency virus type 1 (SF162 isolate) (HIV-1).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90347835; PubMed-2384920;
 RA Cheng-Mayer C., Quiroga M., Tung J.W., Dina D., Levy J.;
 RT "Viral determinants of human immunodeficiency virus type 1 T-cell or
 RT macrophage tropism, cytopathogenicity, and CD4 antigen modulation.";
 RL J. Virol. 64:4390-4398(1990).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M65024; AAA45072.1; -
 DR HIV; M38428; ENVSSP162.
 DR INTERPRO; IPR000228; -
 DR INTERPRO; IPR000777; -
 DR PFAM; PF00516; GP120; 1.
 DR PFAM; PF00517; GP41; 1.
 KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
 KW Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 502 EXTERIOR MEMBRANE GLYCOPROTEIN.
 FT CHAIN 503 847 TRANSMEMBRANE GLYCOPROTEIN.
 FT DISULFID 53 73 BY SIMILARITY.
 FT DISULFID 118 203 BY SIMILARITY.
 FT DISULFID 125 194 BY SIMILARITY.
 FT DISULFID 130 155 BY SIMILARITY.
 FT DISULFID 216 245 BY SIMILARITY.
 FT DISULFID 226 237 BY SIMILARITY.
 FT DISULFID 294 328 BY SIMILARITY.
 FT DISULFID 374 435 BY SIMILARITY.
 FT DISULFID 381 408 BY SIMILARITY.
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 48.9%; Score 2259; DB 1; Length 847;
 Best Local Similarity 49.9%; Pred. No. 2.2e-156;
 Matches 439; Conservative 155; Mismatches 243; Indels 42; Gaps 15;

QY 5 MRAMGRNR-----KLGILYIVMALIIPCLSSQLYATYAGVPVWEDAAVPLFCASDAN 59
 DB 1 MRVKGIRKYNQHLWRGGTLLGLMLI--CSAVEKLWVYVGVVWKEATTLFCASDAK 58

QY 60 LSTKHNWASQACVPTDPTPEHLLVNTDNNFNWENYVQEQEDIIISLWDQSLKPC 119
 DB 59 AYDTEVHNWATHACVPTDPTPEHLLVNTDNNFNWENYVQEQEDIIISLWDQSLKPC 118

QY 120 IQMTFCIMQCNCTDKNNNTSTENRTSSSENPMKTECFNITVTLKDKKQKQALFTVSD 179
 DB 119 VKLTPCVTLCTNLKATNTKSSNWKEMDRGEINCSFKVTTTSIRNKQKAYALFKLD 178

QY 180 LTKLADNTTNTWTLINCNTTQKQPKVSEPIPIYCAPAGYAFICNSAENGTC 239
 DB 179 VVPI-DND--NTSKLINCNTSVITQACPKVSEPIPIYCAPAGYAFICNSAENGTC 235

QY 240 KCSNIVVTCGKPTVSTQILNGTSLKSKIRKMKNSDSGKNIIVTLSDIETCV 299
 DB 236 PCTNVSTVQCTHGRPVVSTQILLNGSLAEGVVRSENFTDNKTIIVOLKESVEINCT 295

QY 300 RPNQNTVOEMKIGP-WMNTYSMALGTGNSRVAICQNTTEWEKALKNTAERYLELNN 358
 DB 296 RP-NNNTRKSTITIGGRAFY--ATGDIIGDIRQAHNCISGEKW---NNTLQIVTKLQA 348

QY 359 TEGNTTIFNRSDGSDVEVTHLHFNCHGEFFVCNTSEMFTFLCNCNNQTSINSA 418
 DB 349 QFGNKTIIVFQKS--SGGDEIVMHSFNGCGEFTFNTQTLNFTW-----NNTIGPNNT 400

QY 419 NGMI--PCKLKQVVRWMMGGGLYAPPIGNLTCTSHITGMLOMDA--PNKNTENFR 474
 DB 401 NGTITLPCRQIINRWQEVGKAMAPPINGQIRCSNITGLLLRDGKEISNTEIFR 460

QY 475 PIGDMKDIWRNELFKYKVVVPFVAPPTIARVIGTGHREKRAVGLMFLGLVLSA 534
 DB 461 PGGDMKDNWRSELKFKYKVVVPFVAPPTIARVIGTGHREKRAVGLMFLGLVLSA 516

QY 535 AGSTMGAATATLAVQVTHSVIKGIVQODNLLRAIQAOQLLSVWGLRQLPARLALET 594
 DB 517 AGSTMGARSLLTVARQLLSGIVQODNLLRAIQAOQLLSVWGLRQLPARLALET 576

QY 595 LIQOQLLNLWCKGRLICYTSVKNNETWRNTNINQIWNGLTWQEDQIDNVSSTIVE 654
 DB 577 YLKOQQLLWCKGRLICYTSVKNNETWRNTNINQIWNGLTWQEDQIDNVSSTIVE 635

QY 655 EIQAQVQEQEKNLELDEWASLWNLDTIKLWYIKIALLIIVGALIGRIVWIVLNL 714

DB 636 LIEESQOQNEQELLELDKQASLWNFIDSKWLIKIFIMVGLVGLRIVFTVLSI 695
 QY 715 VRNTRQGVQPLSLQIPTRQOSEAEPTORTGEGGDEGRPLIPSPQGFLLYTLRLTII 774
 DB 696 VNRVQGYSPLSQTRFPAPRGPDRPEGIEEGGERDRSSPLVHGLLALIWDLRLSLC 755
 QY 775 LMSYHLLSNLISGTQTQVISHRLGLWLTGKIIDACRICAAVHYHWOELQKQATSLIDT 834
 DB 756 LFSYHRLDLILLTAARIVE-----LLGRGWELKYNLWGLWQELKNSAVSLFDA 808

QY 835 FAVAVANWTDIILGILQRLGRGILNIPRRVQGFERSLL 873
 DB 809 IATAVAEGTDRIIEVAQRIAGRAFLHIPRRIRQGFERALL 847

RESULT 8
 ENV_HV1C4
 ID ENV_HV1C4 STANDARD; PRT; 868 AA.
 AC P05879;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
 DE GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].
 GN ENV.
 OS Human Immunodeficiency virus type 1 (CDC-451 isolate) (HIV-1).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87041461; PubMed-3490666;
 RA Desai S.M., Kalyanaram V.S., Casey J.M., Srinivasan A.,
 RA Andersen P.R., Devare S.G.;
 RT "Molecular cloning and primary nucleotide sequence analysis of a
 RT distinct human immunodeficiency virus isolate reveal significant
 RT divergence in its genomic sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8380-8384 (1986).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M13137; AAA44311.1;
 DR PIR; C25233; VCLJH4.
 DR HIV; M13137; ENVSCDC45.
 DR INTERPRO; IPR000328;
 DR INTERPRO; IPR000777;
 DR PFAM; PF00516; GP120; 1.
 DR PFAM; PF00517; GP41; 1.
 DR AIDS; Coats protein; Polypeptide; Glycoprotein; Transmembrane;
 DR Signal.
 FT SIGNAL 1 32
 FT CHAIN 33 522 EXTERIOR MEMBRANE GLYCOPROTEIN.
 FT CDS 523 868 TRANSMEMBRANE GLYCOPROTEIN.
 FT DISULFID 55 75 BY SIMILARITY.
 FT DISULFID 120 216 BY SIMILARITY.
 FT DISULFID 127 207 BY SIMILARITY.
 FT DISULFID 132 163 BY SIMILARITY.
 FT DISULFID 229 258 BY SIMILARITY.
 FT DISULFID 239 250 BY SIMILARITY.
 FT DISULFID 307 341 BY SIMILARITY.
 FT DISULFID 387 456 BY SIMILARITY.
 FT DISULFID 394 429 BY SIMILARITY.
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 48.68; Score 2247.5; DB 1; Length 868;
Best Local Similarity 49.28; Pred. No. 1.6e-155;
Matches 439; Conservative 159; Mismatches 249; Indels 45; Gaps 18;

FT CARBOHYD 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 395 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 828 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 868 AA; 98698 MW; A11527FC52A6F0C8 CRC64;

QY 3 VTRMAGKRRN-----KLGILYIWMALICLSSQLYATYAGVYVDEADAPVFCASD 57
DB 1 MAMKAGIRKNCQLRWGTMGLMLMI--CSAANLVTVYGVYVWKEATVFCASD 58
QY 58 ANLTSTKHNWASQACVPDPTPHEYLNNVDFNFWENYVQEQEDIIISLWDQSLK 117
DB 59 AKAYDEAHNWAHACVPTNPVQVLENFNFNKNWVQEQHEDIIISLWDQSLK 118
QY 118 PCIQMTFMCIONCTDIKNNTSTENRT-----SSSENPMKTCERNITVLDKREK 171
DB 119 PCVKLTPLCVLNTDNTNTTTELSIIVWEQKGMKRNCSNITTSIRDKVORE 178
QY 172 QALFYVSDAKLADN--NTMTLLNCNNTTKQCPKVSFPIPIYCAPAGYAEK 229
DB 179 YALEYKLDVPEIDNKNNTNNTKYLNCNNTSVITQCPKVSFPIPIHYCTPGFALLK 238
QY 230 CNSAFNGTGKCSNISVVTCTHGIKPTVSTQILNGLTSKEIKRMKNISDSKNIIVT 289
DB 239 CNDKKFNGTGCTNVSTVQCTHGIRPVVSTQILLNGSLAEVIRSENFNTNAKTIIV 298
QY 290 LSSDIETCVRPNGNQTQEMKIGP--MAWYSMALGTGS--NRSRVAYCOYNTTEWALK 346
DB 299 LNVSEVINCTRP--NHHTKRRTVTLGPGRVWYV----TGEILGNIRQAHNCISRAQW---N 349
QY 347 NTARYLELINNTGNTMIFNRQDGSDEVTHTLHNGCHGEFFYCNTESEFN--YTFLC 404
DB 350 NTLQIATTLREQGNKTIAFNQ--SGDDPEIVMHSFNCGGEFFYCNSTQLFNSANWVS 408
QY 405 NGT--NCNNTOSFNSANGMIPCKLQVYRSMRGGSLYAPPICGNLTCISHITGMILQMD 463
DB 409 NGTVSVTRKQDGTIIITLPCRIQIINRWQVGVKAMVAPLPIKGLRCSNITGLLTRD 468
QY 464 APW--NKENTFRPGDKMDIWRNELKYYKVVYKVPSPVAPTPIARPVIGTGREKRV 522
DB 469 GGGENQTEIFRPGGDMRDNWSELKYKVKYKIEPLGVAPTAKRRVY----QREKRV 524
QY 523 G-LGMLFLGVLASGTPMGAAATALTQTHSVIKGIVQOQDNLRLATQAOQLRLSWG 581
DB 525 GMLGAMFLGFLGAGSTMGATSMALTQVQROLLSGIVQOQDNLRLATQAOQLRLSWG 584
QY 582 IRLRLARLLALETIQOLLNLWGCKRGLICYTSVKWNETWRNTTINQIWGLNLTQW 641

DB 585 IKQLQARILAVERYLKDQQLLFGWCGSKLICTAVPNASWSNKT-LDQIWNNTWNEW 643
QY 642 DQOQIDNVSSITIEEIQKAOYQOQNEKKLELDEWASLWNLDTKWLYIKIILVGA 701
DB 644 DREIDNTHIITLIEESQOQKQOQELQDKWASLWNTSDITKWLYIKIIFIMVGG 703
QY 702 LIGVRIYMIVLNVRNIROGYQPLSLQIPTRQOQSEATPGRTGGGDEGRPLIPSPQ 761
DB 704 LIGLRIVFAVLGIIVNRVQGYSPLSFOTLLPNRPGDPDPECTEGSGGERGSTRLVHG 763
QY 762 FLPLLYDTRITILWYSYHLLSNLSIGTQTVISHLRLGLWILGOKIIDACRCAAVIHWL 821
DB 764 FLALVMDLRLSLCLFSYHRLRLLLIIVARIVE-----LLGRGWEVLKYWNLLQYWS 816
QY 822 QELQKATSILDTFAVAVANWTDIILGJLGRGILNIPRRVQGFERSLL 873
DB 817 QELKNSAVSLVNVTAIAVAGTDRIEVQYIYRAFLHPRIRQGFERALL 868

RESULT 9
ENV_HV1Y2
ID ENV_HV1Y2 STANDARD; PRT; 843 AA.
AC P35961;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (YU-2 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93021387; PubMed=1404605;
RA Li Y., Hui H., Burgess C.J., Price R.W., Sharp P.M., Hahn B.H.,
Shaw G.M.;
RT "Complete nucleotide sequence, genome organization, and biological
properties of human immunodeficiency virus type 1 in vivo: evidence
for limited defectiveness and complementation.";
RL J. Virol. 66:6587-6600(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M93258; ; NOT_ANNOTATED_CDS.
DR PIR; H44001; H44001.
DR INTERPRO; IPR000328; .
DR INTERPRO; IPR000777; .
DR PFAM; PF00516; GP120; 1.
DR PFAM; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
Signal.
FT SIGNAL 1 29
FT CHAIN 30 489 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 490 843 TRANSMEMBRANE GLYCOPROTEIN.
FT TRANSMEM 738 755 POTENTIAL.
FT DISULFID 53 73 BY SIMILARITY.
FT DISULFID 118 201 BY SIMILARITY.
FT DISULFID 125 192 BY SIMILARITY.
FT DISULFID 130 155 BY SIMILARITY.
FT DISULFID 214 243 BY SIMILARITY.
FT DISULFID 224 235 BY SIMILARITY.
FT DISULFID 292 326 BY SIMILARITY.
FT DISULFID 373 432 BY SIMILARITY.
FT DISULFID 380 405 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).

QY	5	MRAMGRNR-----KLGILYIVMALIIPCLSSSQLYATVIAGVPWEDAAVPLFCASDAN	59
Db	1	MRATEIRKANYOHLWKGGTLLGLMLI--CSAAEQLVTVIYVGVPWKEATTTLFCASDAK	58
QY	60	LTSKEKHNVASQACVPTDTPHXYLLTNVTFNFWNMYVQEQMEDIIISLWDQSLKPC	119
Db	59	AYDTEVHNWATHACVPTDPNPQVLEKLVNTFNFMNKNVQEQMHEDIISLWDQSLKPC	118
QY	120	IOMTFMCIOMNCTDIKNNTSGTENRTSS-----ENPMKTCENITVLKDKKEKQAL	174
Db	119	VKLPLCLVTLNCTDLRN-----ATNTSSWETMEKEIKNCSENIITSLRDVKQKEYAL	173
QY	175	FYVSDLFKLADNNTVNTMTLINCNSTIKQACPKVFPEPIYIYCAPAGYAIFFKNSAE	234
Db	174	FYNLDVVPI-----DNASYRLISCNTSVITQACPKVFPEPIYIYCAPAGFALLKCNDK	228
QY	235	FNGTGCNSIVTCTHGPVSTQLLINGTLLSKEKIRMGKNISDSKNITVLSSDI	294
Db	229	FNGTGPCNTVSTVQCHGIRPVYSTQLLINGSLAEEIVRSFNFTNNKTIIVOLINESV	288
QY	295	EITCVPRGNQTVQEMKIGP-NAMYSMALGTGSGNSRVACQVNTWEKALKNTAEVYL	353
Db	289	VINCTRP--NNTRKSNIGPGRALYTTGEIIGDIRQ--ACNLSLKQWENTLEQITAIKLK	345
QY	354	ELINNTEGNTWLFERSQDGSDEVTHLHFNCHGEFFYCNTESEMYTFLCNGTNCNNTQ	413
Db	346	EQFGN--NKTIIIFPSGGQDPEIYTH--SFNCGGEFFYCNSTQLETW-----NDTR	392
QY	414	SINSA--NGMIPCKLQVVRVSWMRGGSGLYAPPIPGMLTCISHITGMILQMDA--PWKNT	469
Db	393	KLNNTGRNITLPCRILQIINNMQVFGKAMYAPPIRGQIRCSSNITGLLTLTRDGGKDNGT	452
QY	470	ENTFRPIGDMKDIWRNELFYKVKVVRKVPFSVAPTPIARPVIGTGTTHREKRAVGLMFL	529
Db	453	E-IFRPGGDMRDNWRSELYKVKVKEPLGVAPTAKKRVV-----QREKRAVGLGALFL	507
QY	530	GVLSAAGSTMGAATLAVQPHSVKIGVIOQDNLARAQAQELLRLSVGIGIQLRAFL	589
Db	508	GFLGAAGSTMGAASITLVQARQLSSVIOQNNLURAIEAQHLLQITVYWGIGQOLRA	567

RESULT 10

ENV_HV1Z2

ENV_HV	P12487
ENV_HV	P12487

01-OCT
12437

01-OCT

15-JUN-2015

DE
ENVIRONMENTAL
GLYCOPOLYMER
DE

ENV.

OS	Human
DC	Virus

```
NRN
↑↑↑
[1]
```

SEQUENCE

RA Theodor
RL Submit

2000

CC
This S

between the Eu

use b

modifi-

entitled
or sen

U

R EMBL; I
P HTV. M

R HIV; M.
R INTERPI

INTERP

R PEAM; R
R PEAM: R

PFAM; C
AIDS; C

Signal.

SIGNAL
CHATN

CHAIN
CHAIN

DISULFID

DISULFID

DISULFID

DISULFID

DISULFIDISULFIDE

DISULFID

DISULFID

CARBOHY
CARBOHY

CARBOHY
CARBOHYCARBOHY-
CARBOHY-

CARBOHY
CARBOHY

CARBOHY.

10

Query Match 48.5%; Score 2242; DB 1; Length 853;
 Best Local Similarity 49.9%; Pred. No. 3.8e-155;
 Matches 442; Conservative 150; Mismatches 245; Indels 48; Gaps 15;

FT CARBOHYD 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 384 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 608 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 622 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 634 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 671 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 853 AA; 97043 MW; 849B0B8CB8FF7008 CRC64;

40. Query Match 48.5%; Score 2242; DB 1; Length 853;
 Best Local Similarity 49.9%; Pred. No. 3.8e-155;
 Matches 442; Conservative 150; Mismatches 245; Indels 48; Gaps 15;

QY 5 MRANG-KRN-----RKLGLYIVMALIIPCLSSOLYATYAGVPVWEDAAVPLFCASDAN 59
 DB 1 MRVGRICRNCONLKWGIM--LLGILMTCNSADNLWTVYGVVWKEATTLFCASDAK 58
 QY 60 LTSTKHNWASQACVTDPPHPEYLLTNTVDNINENYVQEQWEDIIISLWQSLKPC 119
 DB 59 SYKTEAHNINWATHACVTDPPHPEYLLTNTVDNINENYVQEQWEDIIISLWQSLKPC 118
 QY 120 IOMTFMCIQNMCTD-----IKNNSTSGTENRTSSSENPMKTCFNTTTLVKDKKKQ 172
 DB 119 VKLPVLCVTLNCIDEVMTMKNNV-----TEIRMKNSFNITTVVRDKTKQVH 170
 QY 173 ALFYVSDLTKLADNNTN-TMYTLINCNTTIKQACPKVSEPIPIYCAPAGYAFKCN 231
 DB 171 ALFYRLDIVPIDNDNSTNTYRLNCNTSAITQACPKVSEPIPIHYCAPAGFAILKCR 230
 QY 232 SAENGTGKCSNIVVTGTHGKIKFTVSTOLLNLTLSKEKIRMGKNISDSGKNIIVTLS 291
 DB 231 DKRFNGTGPCTNVSTVQCTHGIRPVSTQLLNGSLAEIEIIRSENLTNAKIIIVQLN 290
 QY 292 SDIEITCVRPNNQTVQEMKIGPMWYSMALGTGNSRSV--AYCOYNTTWEKALNTA 349
 DB 291 ESVAINTCRPYRN-IRORTSIG---LQALYTTKTSIGQALCNISKNENWTLQOVA 345
 QY 350 ERYLELINTTGNMTMIFNRSQDGSDEVTHLHFNCHGEFFYCNTESEMYTFLCNGTNC 409
 DB 346 IKLGNLLN---KTTIIFKPS-SGGDEPITHTSPNGGEPFYCNTESEMYTFLCNGTNC 400
 QY 410 NNTOSINSANGMPCCKLKQVVRSMRGSGLYAPPFGNLTCSHIGTMILQMDAPNKT 469
 DB 401 ANSTESDCKPTTLQCRKIQIINMNGQVGMKAMAPPTEQINCSSNITGLLTRDGGVNS 460
 QY 470 EN-TFRIGGDMKDIWELNFKYKVRKVPFVAPTPTARPVIGTGHREKRAVLGLMLF 528
 DB 461 SNETFERGGDMRNWSELYKVKVLEPLGVAPTAKRRV-----EREKRAVLGLMLF 516
 QY 529 LGVLSAAGSTMGAAATLTQVTHSVIKGIQVQOONLRLAQAQOELLRLSWGIRQLRAR 588
 DB 517 LGFLAAGSTMGARSLLTQVQARLLSGIQQOONLRLAQAQOELLRLSWGIRQLRAR 576
 QY 589 LLALETIQOOLLNNGCKGRILCYISVKWNETWRTNINQIWNLTQWEDQOQIDNV 648
 DB 577 ILAVERYLQOOLLGIGSGKCLICTTTPVWNSWSN-RSLNDIWNQNTWMEWEREDNY 635

QY 649 SSTIYEIQAKVQOQNEKRLLEDEWASLWNLWDLTKWLWYIKIAIIVGALIGVRIV 708
 DB 636 TGLYRLIEESQTOQERNEQELLEDKWASLWNLWDLTKWLWYIKIAIIVGALIGVRIV 695
 QY 709 MIVNLVNRQGYQPLSLQIPTRQQAETPGRTGGGDEGRPLIPSPQGFPLPLYT 768
 DB 696 FAVLSLNRVRQGYSPSLFQTLPPAPRGDPDRGEGGEGGRGSRIRLVNFGFSALIW 755
 QY 769 DLRTIILWSVHLNLSGTOVISHRLGLWILWOKIIDACRIACAAVHVLWLOELOKSA 828
 DB 756 DLRLNCLFSLHRLDLILIAARIVE-----LLGRGWEALKYLNWLQIWSRELKNSA 808
 QY 829 TSLDITFAVANWTDIIILGIGRLGRLNIPRRVQRGFSRL 873
 DB 809 SSLDLTIAVAEGTDRVIEIVRRACRAVLHPTIRIQGLERLL 853

RESULT 11
 ENV_RV153
 ID ENV_RV153 STANDARD; PRT; 852 AA.
 AC P19549;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
 DE GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].
 GN ENV.
 OS Human immunodeficiency virus type 1 (SF33 isolate) (HIV-1).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 RN [1]
 RP MEDLINE-90317906; PubMed-2370688;
 RA York-Higgins D., Cheng-Mayer C., Bauer D., Levy J.A., Dina D.;
 RT "Human immunodeficiency virus type 1 cellular host range,
 RT replication, and cytopathicity are linked to the envelope region of
 RT the viral genome.";
 J. Virol. 64:4016-4020(1990).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M38427; AAA5067.1; -;
 DR HIV; M38427; ENV5SF33.
 DR INTERPRO; IPR000328; -;
 DR INTERPRO; IPR000777; -;
 DR PFAM; PF00516; GP120; 1.
 DR PFAM; PF00517; GP41; 1.
 KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
 KW Signal.
 FT CHAIN 1 29
 FT CHAIN 30 506 EXTERIOR MEMBRANE GLYCOPROTEIN.
 FT CHAIN 507 852 TRANSMEMBRANE GLYCOPROTEIN.
 FT DISULFID 53 73 BY SIMILARITY.
 FT DISULFID 118 206 BY SIMILARITY.
 FT DISULFID 125 197 BY SIMILARITY.
 FT DISULFID 130 156 BY SIMILARITY.
 FT DISULFID 219 248 BY SIMILARITY.
 FT DISULFID 229 240 BY SIMILARITY.
 FT DISULFID 297 331 BY SIMILARITY.
 FT DISULFID 377 439 BY SIMILARITY.
 FT DISULFID 384 412 BY SIMILARITY.
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 608 608 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 808 808 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 848 AA; 96475 MW; 2076751227EC3F3 CRC64;

Query Match 48.4%; Score 2238.5; DB 1; Length 848;
Best Local Similarity 49.0%; Pred. No. 6.8e-155;
Matches 434; Conservative 160; Mismatches 238; Indels 53; Gaps 16;

QY 5 MRAMGKRR-----KLGILYVMAIIPCLSSQLYATVYAGVPEVWEDAAPVLFCSADAN 59
DQ 1 MRVKGIRKNTYQHLWKGGIL--LLGTLMICSAVERKLWTVYGVVWRETTLTFLCASDAK 58

QY 60 LTSTEKINWASQACVTDTPHYELLYTNDNINWENTWVQEQMEDISLWQSKPC 119
DQ 59 AYDEVENWATHACVTPDNPQEVLENTEDFNWKNWVQEQMEDIVNLWQSKPC 118

QY 120 IOWTFMCIOMNCTDIKNNNTSGTENRTSSSENPM-----KTCENITVLKDKKREKQAL 174
DQ 119 VKLPLCVTLNCKDVNATNT-----TSSEGMERGEIKNCSENIKSRDKVQKQYAL 172

QY 175 FYSDLKALADNNTNTMYTLINONSTIKQACPKVSEFPIPIYICAPAGYAIKPCNSAE 234
DQ 173 FYKLDVVP1--DNKNNTKYLISNTSVITQACPKVSEFPIPIYICAPAGYAIKPCNKNT 230

QY 235 FNGTGKSNISVVTCTGIGIPTVSTQILNGLTSKEKIRMGKNISDGNKIIVLSSDI 294
DQ 231 FNGGQCKNISTVQCTGIRPVPVSTQILLNGLSLAEKRVWIRSDNFTDAKTIIVQLNESV 290

QY 295 EITCVRPNNQVQEMKIGP--MAWYSMALGTGSGNRSVAYCOYNTTEWAKLNKTAERYL 353
DQ 291 KINTRESNN--TRKSIHIGPRAFYTTGETIGDIRQ--AHCNISRAQNNWTLKQIVKELR 347

QY 354 ELINTEGNTMTIFNRSQDGSVDVETHLPHNCHGEFFYCNTESEMYTFLCNGTNCNNTQ 413
DQ 348 EQFN---NKTIVTHS--SGGDPEIVMHSFNGCGEYFCNSTQLFNSTW-----NDTE 395

QY 414 SINSANG-----MIPCKLKQVVRSMRGSGLYAPPIGNTICSHITGMILQMDAPNKT 469
DQ 396 KSSGTGNDRIILPCRIKQIINMQEYKAMYPPIKQIRCSSNITGLITRDGGKNES 455

QY 470 E-NTFRPIGDMKDIWENELFKYVVRKVPSPVAPTPIARPVIGTGHREKRAVGLGMLF 528
DQ 456 EIEIFRPGGDMRWSELYKYKVKVIEPLGVAPTAKARKV---QREKRAVIGIGALF 511

QY 529 LGVLSAAGTGAATAALTQVTHSVIKGIVQOQDNLRAIQAOQELLRLSVWGTROLRAR 588
DQ 512 LGFLGAGSTNGARSMTLTQVQARQLLSGIVQOQNNLLRAEQHMLQLTWGLKQLQAR 571

QY 589 LLALETIQOQLNLNCGKRLICYTSVKWNETWRNTNINQIWGLNTWOEQDQIDNV 648
DQ 572 VLAVERYLKQQLMGWICGSGKLCITCTAVPNTWSN--KSLDSIWNNTWMEWEKEIENY 630

QY 649 SSTIYEIOKAOVOEONEKLELDEWASLWNLDTKWLWYIKIATIIIVGALIGRIY 708
DQ 631 TNTIYTLIESQIQOQNEKLELDEWASLWNLDTKWLWYIKIATIIIVGALIGRIY 690

QY 709 MYVLNVRNIRQYQPLSLQIPTRQOSEAETPGTGGGDEGRPRILIPSPQGFPLLYT 768
DQ 691 FSVLSIVNRVRYGYSPLSFQTLTPATRGDREGGERDRDRSGQLVNGFLALIYW 750

QY 769 DLATILWVYHLLSNLSGTTQTVISHRLGLWILGKIIDACRICAAVIVHQLQELOKSA 828
DQ 751 DLKSLFSLFVHRLDLLTTRIVE-----LLGRGWELKYWNWLLQWSELKNSA 803

QY 829 TSLIDTFAVAVANNWTDIIILGIRGLRGIILNPRRVQGFERSLL 873
DQ 804 VSLLNATAIAVAGTDRIIEVQVRYRAILHIPRIQGLERALL 848

RESULT 13
ENV_HV1KB
ID ENV_HV1KB STANDARD; PRT; 861 AA.
AC P31819;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (XB-1 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
CC [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92351552; PubMed-1322587;
RA Shimizu H., Hasebe F., Tsuchie H., Morikawa S., Ushijima H.,
RA Kitamura T.;
RT "Analysis of a human immunodeficiency virus type 1 isolate carrying a
truncated transmembrane glycoprotein.";
RL Virology 189:534-546(1992).
CC -1- MISCELLANEOUS: READING OF THE TERMINATOR UAG OCCURS BETWEEN
THE CODONS FOR 729-ALA AND 730-ARG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DQ ENBL; D12582; BA02124.1; ALT_SEQ.
DQ PIR; A42995; VCLJKB.
DQ PIR; B42995; VCLJKB.
DQ INTERPRO; IPR000328; .
DQ INTERPRO; IPR000777; .
DQ PFAM; PF00516; GP120; 1.
DQ PFAM; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
KW Signal.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 517 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 518 861 TRANSMEMBRANE GLYCOPROTEIN.
FT TRANSMEM 690 711 POTENTIAL.
FT DISULFID 59 79 BY SIMILARITY.
FT DISULFID 124 212 BY SIMILARITY.
FT DISULFID 131 203 BY SIMILARITY.
FT DISULFID 136 160 BY SIMILARITY.
FT DISULFID 225 254 BY SIMILARITY.
FT DISULFID 235 246 BY SIMILARITY.
FT DISULFID 303 337 BY SIMILARITY.
FT DISULFID 383 446 BY SIMILARITY.
FT DISULFID 390 419 BY SIMILARITY.
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 48.3%; Score 2233.5; DB 1; Length 861;
Best Local Similarity 49.8%; Pred. No. 1.6e-154;
Matches 433; Conservative 160; Mismatches 229; Indels 47; Gaps 16;

FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 622 622 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 821 821 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 861 AA; 98116 MW; 3C06787658F0C9DA CRG64;

Query Match 48.3%; Score 2233.5; DB 1; Length 861;
Best Local Similarity 49.8%; Pred. No. 1.6e-154;
Matches 433; Conservative 160; Mismatches 229; Indels 47; Gaps 16;

QY 20 IVMALIICLSSQLXATYAGVVPVWEDAPVLFCSADANLTSTKHNWVASQACVPTDP 79
DB 25 MLGLMICAASQLWYVYGVVKEATITLFCASDAKAYTAHNWATHACVPTDP 84
QY 80 TPHEYLNTVDFNFWNQMEDIIISLDQSLKPCIOFMFCIONMCKDKNNT 139
DB 85 NPOEVVLVNTFNWKNWQMHENIISLDQSLKPCVLTPLCVTLHCTDLRN-- 141
QY 140 SGFENRTSSSENPM---KTCENITVLKDKKEKQALFYVSDTLKL-ADNNTTMT 194
DB 142 --TTNNSSIEEKKGEIKNCSEFNTNDRKVKQYALFVLVDVNDNNSTCYR 199
QY 195 LINCSTTIKQACPKVSFEPIPIYCAPAGYALFKNSAEFNGTKCSNISVVTCHGK 254
DB 200 LISCDTSVITQACPKVSFEPIPIHYCTPAGFALLKCNKNTFNGTGPCKNVSTVQCHGR 259
QY 255 PTYSTQILNGLSKIRIMGNKNSDSGNIIIVTSSDIETCVRPNGNQTVQEMKGP 314
DB 260 PVYSTQLLNGSLAEVGVIRSENFNDVNTIIVQLNETVKINCIRP-NNKTRKRYTMGP 318
QY 315 -MAWYSMALGTGSRNRYVAYQVNTTEWEKALANTAEVLELNNTEGNTMIFNRSQDG 373
DB 319 GRVYTTGELIGDIQO--AHNCISRAEWNKTLQIANK----LRKFENKTVFNS-SG 371
QY 374 SDVEVTHLHFNCHGEFFYCNTESEMYTFLCNGTNCNNTQINSANGMIPCKLKQVVRSW 433
DB 372 GDPEIVMHNFCGGEFFYDSSQLFNSTHLSNGTWNGT---GPNITLPCRICKQIVNMW 428
QY 434 MRGGLYAPPPGNTCTSHITGMILQMDAPWNKENT-----FRPTGGDKMDIWRNE 487
DB 429 QEYGMAYAPPRIQGRCSNITGLLTDGGNTONNTNNSIEIRPFGGDMRDNRSE 488
QY 488 LFYKVVRYKVPFVPTARPIVIGTTHREKRAVGLMFLGLVLSAAGSTWGAATALT 547
DB 489 LYKYVKVIEPLGVAATRAKRRV---OREKRAVIGAVFLGFGAAGSTWGAATVLT 544
QY 548 VQTHSVIKGIVQOQDNLRAIQAOQELLRLSVYGIQRLARLALFELIQOQLNLWGC 607
DB 545 VOARQLLPVQOQDNLRAIQAOQELLRLSVYGIQRLARLAVERYLKDQQLMGWGC 604
QY 608 KGRLCYTSKKNETWNTNINQINGNLTWQEDQOQDINVSSTIYEETQKAOVQEQNE 667

Db 605 SGKFICTTAVPNTWSN-KSFNEIWDNMTNWEEREINNTNLYNLEESQOQEKNE 663
QY 668 KKLLEDEWASLWNLDTIKLWYIKIATIIYVAGLIGVIRVIMVNLVNRINRQGTPLSL 727
Db 664 QULLALDWDLSLWNFSTIKLWYIKIATIIYVAGLIGVIRVIMVNLVNRINRQGTPLSF 723
QY 728 Q--IPTROQSEAEPTGRTGEGGDEGRPLIPSPQGFPLPLTYLTDLTILMSYHLLSN-L 784
Db 724 QTRLPAR--GPDREPTGEEGGERDRSGPLVDGLLALIWDLRLSLCLFSYHRLDRL 780
QY 785 ISGTQTVTSHLGLWILGOKIIDACRCAAVIHWLQELQASLSDTFAVANWTD 844
Db 781 LIVTRTV-----ELLGRKGWEVLKYLWNLQYMSQELKNSAVSLNATAIAAEGTD 832
QY 845 DIIIGTORLGRGILNPRVRGERSLL 873
Db 833 RVIEILQRTYRAILHIPVKIROGLERALL 861

RESULT 14
ENV_HV1W1 STANDARD; PRT; 856 AA.
ID ENV_HV1W1
AC P31872;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (WMJ1 isolate) (HIV-1).
RC Viruses; Retroviridae; Retroviridae; Lentiviridae.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86218077; PubMed=2423250;
RA Starcich B.R., Hann B.H., Shaw G.M., McNeely P.D., Modrow S., Wolf H.,
RA Parks E.S., Parks W.P., Josephs S.F., Gallo R.C., Wong-Skaal F.,
RT "Identification and characterization of conserved and variable
regions in the envelope gene of HTLV-III/LAV, the retrovirus of
AIDS.";
RL Cell 45:637-648(1986).
CC -!- MISCELLANEOUS: ISOLATES WMJ1, WMJ2, AND WMJ3 WERE OBTAINED FROM
BLOOD SAMPLES SEQUENTIALLY TAKEN FROM A TWO-YEAR OLD HAITIAN WHO
WAS PERINATALLY INFECTED BY HER MOTHER.
DR PIR; A24774; VCLJ3W.
DR INTERPRO; IPR000328;
DR INTERPRO; IPR000777;
DR PFAM; PF00516; GP120; 1.
DR PFAM; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
KW Signal.
FT SIGNAL 1 29
FT CHAIN 30 510 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 511 856 TRANSMEMBRANE GLYCOPROTEIN.
FT DISULFID 53 73 BY SIMILARITY.
FT DISULFID 118 205 BY SIMILARITY.
FT DISULFID 125 196 BY SIMILARITY.
FT DISULFID 130 152 BY SIMILARITY.
FT DISULFID 218 247 BY SIMILARITY.
FT DISULFID 228 239 BY SIMILARITY.
FT DISULFID 296 330 BY SIMILARITY.
FT DISULFID 376 444 BY SIMILARITY.
FT DISULFID 383 417 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 276 276 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 384 384 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 442 442 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 612 612 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 812 812 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 852 AA; 97203 MW; 2BB866345DEC915F CRC64;

Query Match 48.2%; Score 2228.5; DB 1; Length 852;
 Best Local Similarity 49.4%; Pred. No. 3.6e-154;
 Matches 442; Conservative 153; Mismatches 229; Indels 71; Gaps 20;

QY 5 MRANG-KRNRK-----LGILYIVMALIIPCLSSQLYATYAGYVWEDAAPVLF 53
 DB 1 MRVGIKKNYQHLRWGGMGLMI-----CSATDKLWTVYVYGVVKEANTLTF 53

QY 54 CASDANLSTKHNWASQACVTDTPHYLELTNTDNFNWENTYVQMOEDIIISLWD 113
 DB 54 CASDAKAYDEIHNWATHACVTPDPNPQELVMGNTNFNMWKNDMVQMHEDIISLWD 113

QY 114 QSLKPCIQMTFMCIQNMCTDIKNNTSGTENSSSE-NPMKTCENITVLKDKKEKQ 172
 DB 114 QSLKPCVKLPLCVTLNCHDF--NATNATNSGKMEGEMKNCSEFNITTSRDMQKEY 171

QY 173 ALFYVSDLIKLANNTNTMYTLINCNSTTIKQACPKVSFEPIPIYCAPAGYAFKNS 232
 DB 172 ALFYKLDIVPI-DNDKTNTRYLISNTSVITQACPVTFEPIPIHYCAPAGFAILKCNN 230

QY 233 AEFNGTKCSNISVVTCTHGKPTSTQILNGLTSKEKIRMGKNISSDGKNIIVTLSS 292
 DB 231 KFEKGTGCTNVTVOCTHGIRPVSTQLLNGSLAEVEVIRSENFNNVNTIIVQLNE 290

QY 293 DIEITVRPNNQTOEMKIGP-MAWYSMALGTGNSRVSAYCOYNTTEWEKALKNTAER 351
 DB 291 SVEINCTRP-NNNTRKRITMGPVYVYTTQIIGDIR--AHCNLSRSKWE----NLIKQ 343

QY 352 YLELINNTEGNTMLFNRSQDSQSDVEVTHLHNCHEFFYNTSEMFNYTFLCNGTNCNN 411
 DB 344 IVTKLRVQKNTIIVFNRS-SGDDPEIYVHNSFCGGEFFCCTQLENSTWY-----RNT 397

QY 412 TOSINSANGMI--PCKLKQVVRSMRGSGLYAPPIGNLTCISHITGMILQMDAPNKT 469
 DB 398 TGNITEGNSPITLPCRIKQIINWQEVGKAMYAPPICQIKCSSNITGLLTRDG-GNNN 456

QY 470 ENT-----FRPIGGDKDWRNELFKYVVRKVPFSAVPTPIARFVIGTGRKRAYGLG 525
 DB 457 ETTDEIFRPGGNNRDNWRSELYKYVVKIEPLGVAPTAKRRV---QREKRAYGLG 512

QY 526 MLFLGLVLSAAGSTMGAAATALTVOHVSIVKIGVOQDNLPAIOAQOELLRLSVGIRQL 585
 DB 513 ALFLGLCAAGSTMGAAATLTQVQARLLSGIVQOQNNLLMAIEAQOHLLETVMGIRQL 572

QY 586 RARLLAETLTONQOLLNMGCKRLCYTSVKWNTRNTNINQWGLTWOEWDQOI 645
 DB 573 QARVLAVERYLKQOQLLIGWCSGLICTTAVPNWASNS-KSLSDIWDNNTWMEWEREI 631

QY 646 DNVSTTIEETQKAOVQOQNEKLLDEWASLWNLDTKWLWYKIAIIVGALIGV 705
 DB 632 DNYTNLIYSLIEDSQIQEKNEKELLEDDKWSLWNNFNITNMLWYKIFIMIVGLIGL 691

QY 706 RIVMIVLNVRNIROGYOPLSLQIPTROQSEAEPTGRTG-----EGGDEGPRPLIP 757
 DB 692 RIVFAVLSIVNRVROGYSPLSFQ-----TRLPGRGPDPRPEIEEGGERDRDRSSP 743

QY 758 SPQGFPLLYTDLRITIIILWSYHLLSNLISGQTQTVISHRLGLWILGKIIDACRICAIVI 817
 DB 744 LYDGFALFWVDLRSFLFSYHRLRDLILLIIVTRIVE-----LLGRGWEVLKYWNLL 796

QY 818 HYWLQELQKSAFSLIDTFAVAVANWNTDDIILGQIPLGRGILNIPRVQGFERSL 872
 DB 797 QIWSQELKNSAVSLUNATAIAGERTDRAIEVQVAFRAILHIPRIRIQGLERAL 851

Search completed: January 25, 2001, 11:02:20
 Job time: 655 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2001, 10:51:24 ; Search time 15.24 Seconds
(without alignments)
1028.642 Million cell updates/sec

Title: US-08-911-824-61
Perfect score: 4622
Sequence: 1 MIVTMRANGKRNKLGILYI.....GKILNIPRRVQGFERSLL 873

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/3A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/3B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/PTUS_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4622	100.0	873	2	US-08-912-129A-61
2	2250	48.7	855	1	US-08-022-835-6
3	2250	48.7	855	1	US-08-388-809-6
4	2250	48.7	855	2	US-08-647-714-6
5	2248	48.6	857	2	US-08-448-603A-30
6	2248	48.6	857	3	US-09-134-075-30
7	2247	48.6	850	2	US-08-448-603A-28
8	2247	48.6	850	3	US-09-134-075-28
9	2211.5	47.8	856	1	US-08-022-835-2
10	2211.5	47.8	856	1	US-08-388-809-2
11	2211.5	47.8	856	2	US-08-647-714-2
12	2211.5	47.8	863	3	US-08-463-210-11
13	2211	47.8	857	1	US-08-022-835-4
14	2211	47.8	857	1	US-08-388-809-4
15	2211	47.8	857	2	US-08-647-714-4
16	2208.5	47.8	856	2	US-07-916-098A-2
17	2205.5	47.7	861	1	US-08-127-499A-14
18	2205.5	47.7	861	1	US-08-482-847-14
19	2190.5	47.4	880	3	US-08-788-815-7
20	2190.5	47.4	880	3	US-09-157-963-7
21	2181.5	47.2	826	1	US-08-375-510-2
22	2181.5	47.2	826	1	US-08-487-657-2
23	2175.5	47.1	856	1	US-08-375-100-1
24	1926.5	41.7	618	2	US-08-912-129A-54
25	1921	41.6	373	2	US-08-912-129A-52
26	1791.5	38.8	665	2	US-08-448-603A-32
27	1791.5	38.8	665	2	US-09-134-075-32
28	1682	36.4	600	2	US-08-392-806A-4

29	1678	36.3	606	2	US-08-392-806A-6	Sequence 6, Appl
30	1657	35.9	596	2	US-08-392-806A-2	Sequence 2, Appl
31	1452	31.4	519	1	US-08-589-446-8	Sequence 8, Appl
32	1452	31.4	519	1	US-08-444-882-8	Sequence 8, Appl
33	1452	31.4	519	2	US-08-389-459A-8	Sequence 8, Appl
34	1452	31.4	519	3	US-08-987-867A-8	Sequence 8, Appl
35	1428.5	30.9	351	1	US-08-470-202-46	Sequence 46, Appl
36	1428.5	30.9	351	1	US-08-471-770-46	Sequence 46, Appl
37	1428.5	30.9	351	2	US-08-468-059-46	Sequence 46, Appl
38	1346	29.1	356	1	US-08-602-713-12	Sequence 12, Appl
39	1346	29.1	356	3	US-08-989-493-12	Sequence 12, Appl
40	1316	28.5	494	3	US-08-889-841-19	Sequence 19, Appl
41	1314.5	28.4	495	3	US-08-889-841-28	Sequence 28, Appl
42	1308.5	28.3	474	3	US-08-889-841-36	Sequence 36, Appl
43	1308.5	28.3	474	3	US-08-889-841-39	Sequence 39, Appl
44	1308.5	28.3	495	3	US-08-889-841-25	Sequence 25, Appl
45	1305.5	28.2	486	3	US-08-889-841-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-912-129A-61
; Sequence 61, Application US/08912129A
; Patent No 5922533
; GENERAL INFORMATION:
; APPLICANT: VALLARI, ANADRUZELA S.
; APPLICANT: HACKITT, JOHN JR.
; APPLICANT: HICKMAN, ROBERT K.
; APPLICANT: VARIATEK, VINCENT A. JR.
; APPLICANT: NECKLAWS, ELIZABETH A.
; APPLICANT: GOLDEN, ALAN M.
; APPLICANT: BRENNAN, CATHERINE A.
; APPLICANT: DEVARIE, SUSHIL G.
; TITLE OF INVENTION: RAPID ASSAY FOR SIMULTANEOUS DETECTION AND DIFFERENTIATION
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.44 MB
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS (Windows 95)
; SOFTWARE: Microsoft Word (ASCII format output)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/912,129A
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Danckerts, Andreas M.
; REGISTRATION NUMBER: 32,652
; REFERENCE/DOCKET NUMBER: 6109.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-937-9803
; TELEFAX: 847-538-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 873 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-912-129A-61

```

Query Match      100.0%; Score 4622; DB 2; Length 873;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYTMRAMGRNKLGLIYVMAIIICLSSSOLYATYAGVPVWEDAAAPVLFCSADANL 60
DB 1 MYTMRAMGRNKLGLIYVMAIIICLSSSOLYATYAGVPVWEDAAAPVLFCSADANL 60
QY 61 TSTEKNHNVASQACVPTDPTPHEVLLTNTDNIENWYVQMQEDIIISLWQSLKPCI 120
DB 61 TSTEKNHNVASQACVPTDPTPHEVLLTNTDNIENWYVQMQEDIIISLWQSLKPCI 120
QY 121 QMTFMCIONMCTDIKNNTSGTENRTSSSENPMKTCFENITVLDKKEKQALFYVSD 180
DB 121 QMTFMCIONMCTDIKNNTSGTENRTSSSENPMKTCFENITVLDKKEKQALFYVSD 180
QY 181 TKLADNNTNTMYTLNCHSTIKQACPKVSEPIPIYCAPAGYAFKCSAEFNGTK 240
DB 181 TKLADNNTNTMYTLNCHSTIKQACPKVSEPIPIYCAPAGYAFKCSAEFNGTK 240
QY 241 CSNISVVTCTHGIKPTVSTQILNGLTSLKEKIRIMGNISDSGKNIIVTSLSDIEITCVR 300
DB 241 CSNISVVTCTHGIKPTVSTQILNGLTSLKEKIRIMGNISDSGKNIIVTSLSDIEITCVR 300
QY 301 PGNNQTVQEMKIGPMWNTSMALGTGNSRNVAYCOYNTTEWEKALNTAERYLELNTE 360
DB 301 PGNNQTVQEMKIGPMWNTSMALGTGNSRNVAYCOYNTTEWEKALNTAERYLELNTE 360
QY 361 GNTTMIENRSDGSDVEVTHLFHNCHEFFCYNTSEMNTFLCNGTNCNNTQINSANG 420
DB 361 GNTTMIENRSDGSDVEVTHLFHNCHEFFCYNTSEMNTFLCNGTNCNNTQINSANG 420
QY 421 MIPCKLKQVRSWMRGSGLYAPPIGNLTCTSHITGMILQMDAPWKNKTENFRPIGDM 480
DB 421 MIPCKLKQVRSWMRGSGLYAPPIGNLTCTSHITGMILQMDAPWKNKTENFRPIGDM 480
QY 481 KDIWNEFLFKYVVRKPFVAPPIAPVPIGTGTHREKRAVGLGMLFLGVLAAAGSTM 540
DB 481 KDIWNEFLFKYVVRKPFVAPPIAPVPIGTGTHREKRAVGLGMLFLGVLAAAGSTM 540
QY 541 AAATALTAVQTHSVIKIVQOQDLRAIQAEQLRLSVGIRQLRALLALETILQNOQ 600
DB 541 AAATALTAVQTHSVIKIVQOQDLRAIQAEQLRLSVGIRQLRALLALETILQNOQ 600
QY 601 LLNLWGCKGRILCYTSVKWNETWNTNTNINQWNLFWQEDWQOQIDNVSTIYEEIQKAQ 660
DB 601 LLNLWGCKGRILCYTSVKWNETWNTNTNINQWNLFWQEDWQOQIDNVSTIYEEIQKAQ 660
QY 661 VQOEQNEKKLELDEWASLWNLIDITKWLVIKIAIIVGALIGRIVMVLNVRNIQ 720
DB 661 VQOEQNEKKLELDEWASLWNLIDITKWLVIKIAIIVGALIGRIVMVLNVRNIQ 720
QY 721 GYQPLSLQIPTROQSEAEPTGRTGEGGDEGRPLIPSPQGLPLLYTLDRITLWSYHL 780
DB 721 GYQPLSLQIPTROQSEAEPTGRTGEGGDEGRPLIPSPQGLPLLYTLDRITLWSYHL 780
QY 781 LSNLSIGTORVISHRLGLWILGOKIIDACRICAAVHIVWLOKASATSLIDTFAVAVA 840
DB 781 LSNLSIGTORVISHRLGLWILGOKIIDACRICAAVHIVWLOKASATSLIDTFAVAVA 840
QY 841 NWTDDIILGIRGLRILNIPRRVQGFERSLL 873
DB 841 NWTDDIILGIRGLRILNIPRRVQGFERSLL 873

```

RESULT 2

US-08-022-835-6
; Sequence 6, Application US/08022835
; Patent No. 5420030
; GENERAL INFORMATION:
; APPLICANT: Reitz Jr., Marvin S.
; APPLICANT: Franchini, Genoveffa

```

; APPLICANT: Markham, Phillip D.
; APPLICANT: Gallo, Robert C.
; APPLICANT: Loxi, Franco C.
; APPLICANT: Popovic, Mikulas
; APPLICANT: Garter, Suzanne
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: Eleventh Floor, 1615 L. Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/022,835
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/599,491
; FILING DATE: 17-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Scott, Watson T.
; REGISTRATION NUMBER: 26,581
; REFERENCE/DOCKET NUMBER: WTS/5683/78507/SRL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEFAX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 855 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-022-835-6

```

Query Match 48.7%; Score 2250; DB 1; Length 855;
Best Local Similarity 51.6%; Pred. No. 2e-176;
Matches 448; Conservative 138; Mismatches 239; Indels 44; Gaps 18;

```

QY 15 LGILYIVMALIIPCLSSSOLYATYAGVPVWEDAAAPVLFCSADANLSTTEKHNVASQAC 74
DB 21 LGILMI-----CNAEKLWTVYVYGVVWKEATITLFCASDRKAYDTEVHNWATHAC 73
QY 75 VPTDPTPHEVLLTNTDNIENWYVQMQEDIIISLWQSLKPCIOMTFCIONMCTDI 134
DB 74 VPTDPTPHEVLLTNTDNIENWYVQMQEDIIISLWQSLKPCIOMTFCIONMCTDI 133
QY 135 KNNNTSGTENRTSSSE-----NPMKTCFENITVLDKKEKQALFYVSDTLKLANNTT 189
DB 134 RNATNGNDTNTSSRGWGGGKNCNFNTTNGKVKQKALFYKLDIAPIDNN-S 191
QY 190 NMTYTLNCHSTIKQACPKVSEPIPIYCAPAGYAFKCSAEFNGTKCSNISVWTC 249
DB 192 NNRYLISCVITVITACPKVSEPIPIHYCAPAGYAFKCSAEFNGTKCSNISVWTC 251
QY 250 THGKPTVSTQILNGLTSLKEKIRIMGNISDSGKNIIVTSLSDIEITCVRGNNNTVOE 309
DB 252 THGIRPVSTQILNGLTSLKEKIRIMGNISDSGKNIIVTSLSDIEITCVRGNNNTVOE 310
QY 310 MKTIGP-MANYSMALGTGNSRNVAYCOYNTTEWEKALNTAERYLELNTEGNTMIEN 368
DB 311 IHIGPGRAFTTGEIIGDIRQ--AHCNLSRAKWNDFLNKIVIKLREQF----GNKTIVFK 364
QY 369 RSDGSDVEVTHLFHNCHEFFCYNTSEMNTFLCNGTNCNNTQINSANGMIPCKLK 427

```


Db 365 HSSGDDPEIVTH-SFNCGGDEFFCYNSTOLFENSTW--NVTEESNNVTENNIT--LPCRIK 419
QY 428 QVSRWNRGSGGLYAPPPIGNLTICSHITGMILQMDA--PWNTENTFRIGGDMKDIWR 485
Db 420 QIINWQEVGRAMYAPPIRGICRSCSNITGLLTRDGGPEDNKE-VFRPGGDMRDNR 478
QY 486 NELFKYKVRVKPSPVAPTPIARPVIGTGTTHREKRAVGLGMLFLGVLSAAGSTMGAATA 545
Db 479 SELYKYKVKVIEPLGVAPTAKRRVV---QREKRAVGIGAVFLGFLGAAGSTMGAATA 534
QY 546 LTQVTHSVKIGVQQDNLRAIOAOQELLRLSWGIRQLRALLALETILQOQLNLW 605
Db 535 LTQVALLSGVQQDNLRAIOAOQELLRLSWGIRQLRALLALETILQOQLNLW 594
QY 606 GCKGRLLCYTSVKWNETWNTNINQWGLTWQEDQIDNVSTIIEIQAQVQOE 665
Db 595 GCSGLKLTAVPWNASWSN-KSLNKIWDNTWIEWDRINNTSYIISLESQNOQEK 653
QY 666 NEKKLELDEWASLWNLDTIKWLYIKIAIIVGALIGVIRVIMVLNLRNRQGYOPL 725
Db 654 NEQELLELDKWSLWNLDTIKWLYIKIIFIMIVGGLIGLRIVSVLSIVNRVQGYOPL 713
QY 726 SLQIPTRQOSEAETPGTGGGDEGRPLIPSPQGFPLLYTDLRTIILWSHLLSNLI 785
Db 714 SFQTHLPSSRGDPDRPGGIEEGGERDRSGPLVNGFLALIWDLRSLFLSFYHRLDLL 773
QY 786 SGTQTVISHURL-GLWILGOKIDACRICAATHYIQLQKSAATSLIDTFAVANWTD 844
Db 774 LIVNRIVELLGLAGW-----EVLKYNWNLQYWSQELKNSAVSLINATAVABEGTD 826
QY 845 DIILGIQRLGRGILNIPRRVROGERSLL 873

RESULT 3

US-08-388-809-6
; Sequence 6, Application US/08388809
; Patent No. 5576000
; GENERAL INFORMATION:
; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,
; APPLICANT: GENOVEFFA, MARKHAM, PHILIP D. GALLO, ROBERT
; APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
; APPLICANT: GARTNER, SUZANNE
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK, 3.5"
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/388,809
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,835
; FILING DATE: 25-FEB-1993
; APPLICATION NUMBER: US 07/599,491
; FILING DATE: 17-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: LESLIE A. SERUNIAN
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4092052
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 855 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-388-809-6

Query Match 48.7%; Score 2250; DB 1; Length 855;
Best Local Similarity 51.6%; Pred. No. 2e-176;
Matches 448; Conservative 138; Mismatches 239; Indels 44; Gaps 18;

QY 15 LGILYIVMALIIFCLSSQLYAVYAGVPWDEAAPVLFCAASDANLTSTKHNWASQAC 74
Db 21 LGILMI-----CNAEEKLWVTYVGVVWKEATTFLFCASDRKAYDTEVHNWATHAC 73
QY 75 VPTDPTPHEYLNTDNFNINWVQMOEDIISLWQSLKPCIQMTFCOMNCTDI 134
Db 74 VPTDPNPQEVLEKNVTENFNWKNWVQMHEDIISLWQSLKPCVCLTFLCYLNTCLDL 133
QY 135 KNNISGTEENRTSSSE-----NPMKTCFENITTVLKKKQALFYVSDLTKLADNNT 189
Db 134 RNATNGNDTNTSSRGVGGGEMKCNCFNITNIRGVQKEVALFYKLDIAPIDNN-S 191
QY 190 NMYTLINCNSITKQACPKVSEPIYCAPAGYAFKCNSEAFNGTKCSNISVVC 249
Db 192 NNRYLISCNTSVITQACPKVSEPIYHICAPAGFAILCKDKKFGKPGCTNVSTVQC 251
QY 250 THGIKPTVSTQILNGTLSEKIRIMGKNSDSGKNIIVTLSSDIETCVRPGNNOTVQE 309
Db 252 THGIRPVSTQILLNGSLAEVEVIRSANFADNAKVIIVQLNESVEINCTRP-NNNTRKS 310
QY 310 MKIGP-MANYSMALGTGSGNRSVAYCOYNTTEWEKALKNTAERYLELINNTEGNTMIFN 368
Db 311 IHIGPGRAFYTTEIIGDIRQ---AHCNLSRAKWNNTLKNIVIKLREQF---GNKTIVFK 364
QY 369 RSQDSDVEVTHLHFNCHGEFFCYNTSEMYNIFLCNGT-NCNNTOSINSANGMIFCKLK 427
Db 365 HSSGDDPEIVTH-SFNCGGDEFFCYNSTOLFENSTW--NVTEESNNVTENNIT--LPCRIK 419
QY 428 QVSRWNRGSGGLYAPPPIGNLTICSHITGMILQMDA--PWNTENTFRIGGDMKDIWR 485
Db 420 QIINWQEVGRAMYAPPIRGICRSCSNITGLLTRDGGPEDNKE-VFRPGGDMRDNR 478
QY 486 NELFKYKVRVKPSPVAPTPIARPVIGTGTTHREKRAVGLGMLFLGVLSAAGSTMGAATA 545
Db 479 SELYKYKVKVIEPLGVAPTAKRRVV---QREKRAVGIGAVFLGFLGAAGSTMGAATA 534
QY 546 LTQVTHSVKIGVQQDNLRAIOAOQELLRLSWGIRQLRALLALETILQOQLNLW 605
Db 535 LTQVALLSGVQQDNLRAIOAOQELLRLSWGIRQLRALLALETILQOQLNLW 594
QY 606 GCKGRLLCYTSVKWNETWNTNINQWGLTWQEDQIDNVSTIIEIQAQVQOE 665
Db 595 GCSGLKLTAVPWNASWSN-KSLNKIWDNTWIEWDRINNTSYIISLESQNOQEK 653
QY 666 NEKKLELDEWASLWNLDTIKWLYIKIAIIVGALIGVIRVIMVLNLRNRQGYOPL 725
Db 654 NEQELLELDKWSLWNLDTIKWLYIKIIFIMIVGGLIGLRIVSVLSIVNRVQGYOPL 713
QY 726 SLQIPTRQOSEAETPGTGGGDEGRPLIPSPQGFPLLYTDLRTIILWSHLLSNLI 785
Db 714 SFQTHLPSSRGDPDRPGGIEEGGERDRSGPLVNGFLALIWDLRSLFLSFYHRLDLL 773
QY 786 SGTQTVISHURL-GLWILGOKIDACRICAATHYIQLQKSAATSLIDTFAVANWTD 844
Db 774 LIVNRIVELLGLAGW-----EVLKYNWNLQYWSQELKNSAVSLINATAVABEGTD 826
QY 845 DIILGIQRLGRGILNIPRRVROGERSLL 873


```

Db 174 FNKLDIVPDDRRSTRNSTRNRYLSCNTSVITQACPKVSFEPIPHCTPAGFALLKC 233
Qy 231 NSAENFGTKSNTSVVCTHGIKPTVSTQILNGTISKIRIMGNISDSKNIIVTL 290
Db 234 NKKFNGSGPKNVSTVOCTHGIKPTVSTQILNGTISKIRIMGNISDSKNIIVTL 293
Qy 291 SSDIEITVRPFGNQTQVEMKIGP-MAWYSMALGTGSRNRVAYCOVNTWEKALKNTA 349
Db 294 TEPVKINTRP-NNTRKSPIGPGAFY--ATGDIIGNRQAHCHNLSRDM---NNTL 346
Qy 350 ERYLELNNTEGNTMIFNPSQDSGVVEVTHLHFNCHGEFFYCNTESEMYTF- 405
Db 347 GQIVKLEQFGNKTIFNHS-SGGDEIYVHFNCHGEFFYCNTESEMYTF- 405
Qy 406 GTNCCNTQSNISANGMIPCKLKOVRSWRRGSGGLVAPPIPGNLCTISHTGMILQDAP 465
Db 406 GTSTEENSTIT---LPCRIKOIVNMWQEVGKAMYPPIRGQIRCSNITGLLTRDGG 460
Qy 466 WNKTEN-TFRPIGDMKDINRNELFKYVYVVPFVSAPTPIARPVIGTGTREKRAVGL 524
Db 461 SNNSMNETFRPGGDMRDNWRSELYKVKVIEPLGVAPTAKRRVY---QREKRAVGI 516
Qy 525 GMLFGLVLSAAGSTMGAATATVQVHSVIGKIVQOQDMLLRAIQAOQELLRLSWGIRQ 584
Db 517 GAVFLGFLGAGSTMGAASTLTQVQARLLSGIVQOQDMLLRAIQAOQELLRLSWGIRQ 576
Qy 585 LRARLALLETIONQOLLNMGCKGLICYSVKNWETWRNTNINQWGLNLTQWEDQO 644
Db 577 LQARVLAVERYLQDQLLWGCSSGLICTTSPVNASWSN-KSLDKIWDNNTWMEWRE 635
Qy 645 IDNVSSTIYEETQKAOQVQOQNEKLELDEWASLWNNLDTIKLWYIKIAIIVGALIG 704
Db 636 IENYTLITLLEESQOQNEKLELDEWASLWNNLDTIKLWYIKIAIIVGALIG 695
Qy 705 VRIVMVLNVRNIRQGOPLSLQIPTROQSEAPTPGRTGGGDEGRPLIPQGFPLP 764
Db 696 LRIVEAVLSVNRVROGISPLSFQRLPAPRPRDPEIEEGEGQGRDRSIRLVQDGLA 755
Qy 765 LXTDLRTILWYHLSNLISGTQVISHLRLGLWILGKIIDACRICAIAVHWIQL 824
Db 756 LWDOLRLSLCFSYHRLDLIIATRIYE-----LLGRGWALKYWNLLQYWIQL 808
Qy 825 QKSATSLDITFAVANVTDDIILGQIRGLRGLNIPRVQGFERSIL 873
Db 809 KNSAVSLNNTAIAVAGTDRVLEVLQRAYRAILHPIRIRQGLERALL 857

```

RESULT 7

```

US-08-448-603A-28
; Sequence 28, Application US/08448603A
; Patent No. 5864027
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; APPLICANT: Nakamura, Gerald R.
; TITLE OF INVENTION: HIV Envelope Polypeptides
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McCutchen, Doyle, Brown & Enersen, LLP
; STREET: 3 Embarcadero Center
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; * MEDIUM TYPE: Diskette
; * OPERATING SYSTEM: DOS
; * SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,603A
; FILING DATE: 07-JUN-1994
; CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/072,833
; FILING DATE: 07-JUN-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Halliday, Emily
; REGISTRATION NUMBER: 38903
; REFERENCE/DOCKET NUMBER: 14918-704
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-393-2000
; TELEFAX: 415-393-2286
; TELEX:
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 850 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-448-603A-28

```

```

Query Match 48.6%; Score 2247; DB 2: Length 850;
Best Local Similarity 50.2%; Pred. No. 3.5e-176;
Matches 432; Conservative 157; Mismatches 236; Indels 36; Gaps 14;

```

```

Qy 20 IYVALLIPCLSSQLYATYVYVVPWEDAAPVLFASDANLSTSEKHNYWASQACVPTDP 79
Db 19 MLLGLMLICSAEKLWTVYVYVVPWKEATTLFCASDAKAYDTEVHNWATHACVPTDP 78
Qy 80 TPHEYLITNTVDFNFWYEQEDIIISLWDSQSLKPCIOQWFCIQMCTDIKN--N 137
Db 79 NPQIEIGLVNTFNWKNWYEQMHEDIISLWDSQSLKPCVLTPLCVTLNCTDKNATN 138
Qy 138 NTSGETNRTSSSENPMKTCENITVLKDKKEKQALFYVSDLTKLADNNTNTWTNLTIN 197
Db 139 TTSSGKMKMERGE--IKNSFNWTSIRDKMNEVALFYKLDVVP1-DND--NTSYRLIS 193
Qy 198 CNSTTLKQACPKVSFEPIIYCAPAGYALFKCSAEENGTKCSNISVVTCTHGIKPTV 257
Db 194 CNTSVITQACPKVSFEPIIYCAPAGFAILLKCRDKFNGTGPCTNVSTVQCTHGIKPTV 253
Qy 258 STQILNGTSLSEKIRIMGNISDSKNIIVTSLSDIEITCVRPGNNTVQEMKIGP-MA 316
Db 254 STQILLNGSLAEEVIRSANFSDNAKTIIVQLNESVEINCRP--NNTRSIHIGPRA 312
Qy 317 WYSMALGTGSRNRVAYCOVNTWEKALKNTAERYLELNNTEGNTMIFNRSQDSGV 376
Db 313 FY--ATGEIIGDIRQAHCHNLSSTKNWNTLKOIVTKLREHFNK-----TIVFNHS--SGGDP 364
Qy 377 EYTHLHFNCHGEFFYCNTESEMF----NYTFLCNGTNCNNTQINSANGMIPCKLKOVVRS 432
Db 365 EIVMHSFNGCGEFFYCNTPLFNSTWNTYTNWNTGESND---TGRNITLQCRKIQIINM 421
Qy 433 WMRGSGLYAPPIPGNLCTISHTGMILQDAPWNKTENTFRPIGGDMKDIWRNELFKYK 492
Db 422 WQEVGKAMYPPIRGQIRCSNITGLLTRDGGNNSSETIFRPGGDMRDNWRSELYK 481
Qy 493 VYRVKPFSAVPTPIARPVIGTGTREKRAVCLGMLFGLVLSAAGSTMGAAATALTQVTHS 552
Db 482 VYKIEPLGVAPTAKRRVY---QREKRAVIGAVFGLGAAGSTMGAAASVTLTVQARL 537
Qy 553 VIKGIVQOQDMLLRAIQAOQELLRLSWGIRQLRALLLETILQOQLNLWGCGRLLI 612
Db 538 LLSGIVQOQDMLLRAIEAQHLLQTLVWIKQLQARVLAVERYLKQOQLLWGCSSGLI 597
Qy 613 CYTSVKKNWETWRNTNINQWGLNLTQWEDQOQDNDVSSYIIEEIQKAOVQOQNEKLE 672
Db 598 CTAVPNNASWSN-KSLDKIWDNNTWMEWREIDNTSLIYSIESQOQNEKLE 656
Qy 673 LDWASLWNNLDTIKLWYIKIAIIVGALIGVIRVIMVLNVRNIRQGOPLSLQIPTR 732
Db 657 LDKWASLWNNLDTIKLWYIKIFIMVGLVGLVIRVITVLSIVNRVRYKGYSLSFOTHP 716
Qy 733 QQSEAEPTGRTGGGDEGRPLIPSPQGFPLLYTLDTLRTIILWYHLSNLISGTQV 792

```

Db 717 APRGLDRPEGEGERDRSRRLVDGFLAIYVWDLRSCLFSYHRLRDLIIAARIY 776
QY 793 SHRLGLWLGOKIIDACRICAAVHYWLOKQSAATSLDITFAVAVANTDDIILGIOR 852
Db 777 E-----LLGRGWELKAYWNNLQYWIQELKNSAVSLNATAIAVAEGTRVIEIVQR 829
QY 853 LGRGILNIPRRVROGFERSLL 873
Db 830 AYRAILHIPTRIOGLERALL 850

RESULT 8
US-09-134-075-28
; Sequence 28, Application US/09134075
; Patent No. 6042836
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; APPLICANT: Nakamura, Gerald R.
; TITLE OF INVENTION: HIV Envelope Polypeptides
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McCutchen, Doyle, Brown & Ebersen, LLP
; STREET: 3 Embarcadero Center
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/448,603
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haliday, Emily
; REGISTRATION NUMBER: 38903
; REFERENCE/DOCKET NUMBER: 14918-704
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-393-2000
; TELEFAX: 415-393-2286
; TELEX:
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 850 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-134-075-28

Query Match 48.68; Score 2247; DB 3; Length 850;
Best Local Similarity 50.28; Pred. No. 3.5e-176;
Matches 432; Conservative 157; Mismatches 236; Indels 36; Gaps 14;

QY 20 IVMAIIPCLSSQLYATYVAGVPWEDAAPVLFASDANLTSTKEKNVWASQACVPTDP 79
Db 19 MLLGMLMCSAAELWTVYGVFWKEATITLFCASDAKAYDEVINWATHACVPTDP 78
QY 80 TPHEYLNTVNTNINVENYVMEQMEDIIISLWQSLKPCIQMTFMCIONCTDIKN--N 137
Db 79 NPQSIGLVNTFNENMKNVMEQMHEDIISLWQSLKPCVKLPCLVTLNCTDLKNATN 138
QY 138 NTSGTENTSSSNPMKTCFNITVILKDKKQALFYVSDILKLDANNNTNTMYTLIN 197
Db 139 TTSSSGMKMERGE--IKNCSFNVTISDRMKNEALFYKLDVVYPI-DND--NTSYRLIS 193

QY 198 CNSTTIQACPKYVFPPIYIYCAPAGYALFKCNSEAFNGTKSCNSIVVYTCFHGKPTV 257
Db 194 CNTSVITQACPKYVFPPIYIYCAPAGFALLKCRDKKFNKGTPCTNVSTVQCTHGRPVV 253
QY 258 STOLLINGTLSEKIRIMGNISDSGNIIIVTLSSDIEITCVPRGNQTVQEMKIGP-MA 316
Db 254 STOLLINGSAAFEVIRSANFSDNAKTIIVQLNESVEINCTRP--NNNTRRSIHGPGRA 312
QY 317 WYSMALGTGNSRSRVAYQYNTTEWEKALKNTAEVLELINTEGNTMTLNFNSQDSDV 376
Db 313 FY--ATGEIIGDIRQAHNLSSTKNWNTLQIVTKLREHFK-----TIVFNIS--SGDP 364
QY 377 EVTHLFNCHGEFFYCNTESEMF-----NYTFLCNGTNCNNTQNSANGMIPCKLKQVRS 432
Db 365 EIVMHSFNGCGEFCYNTTFLFNSNTWYTVTWNTEGSD--TGRNITLQCRIKQIINM 421
QY 433 WMRGSGLYAPPINLTCLSHITGMLQMDAPWNTKENTFRPIGDMKDIDWNEFLPKYK 492
Db 422 WQEVGKAMYAPPIRGQIRCSNITGLLTRDGGNNSETEIFRPGGDMRDNWRSLEYKYK 481
QY 493 VVRVKPFSVAPTPIARPVIGTGHREKRAYGLGMLFLGVLSSAAGSTGMAAATALTVOH 552
Db 482 VKIEPLGVAPTAKRRVM-----QREKRAYGIGAVFLGLAAGSTGMAASVTLTVQARL 537
QY 553 VIRGIQQQDNLRAIQAOQELRLSVGIRQLARLALLETILQOQLLNLMLGCKGRLI 612
Db 538 LLSSIVQQNNLLRAIEAQHLLQLTWGIKQLQARVLAVERYLKDQQLLGIWGCCKLI 597
QY 613 CYTSVKNWETWNTNINQWNLTWQEDQOQIDNVSSITVEIQRAVOQOQNEKKLLE 672
Db 598 CTTAVPWNASWSN--KSLDKIWNNTWMEWEIDNYSITSLIYESQNOQKNEQELLE 656
QY 673 LDEWASLWNLDTTKLWYIKIATIIIVGALIGVIRVIMVLNVRNIROGYOPLSLQIPTR 732
Db 657 LDKWASLWNLDTTKLWYIKIATIIIVGALIGVIRVIMVLNVRNIROGYOPLSLQIPTR 716
QY 733 QQSEAEPTGRTGEGGDEGRPLIPSPQGFPLLYTLDTLTIILWSYHLLSNLSGTQTVI 792
Db 717 APRGLDRPEGEGERDRSRRLVDGFLAIYVWDLRSCLFSYHRLRDLIIAARIY 776
QY 793 SHRLGLWLGOKIIDACRICAAVHYWLOKQSAATSLDITFAVAVANTDDIILGIOR 852
Db 777 E-----LLGRGWELKAYWNNLQYWIQELKNSAVSLNATAIAVAEGTRVIEIVQR 829
QY 853 LGRGILNIPRRVROGFERSLL 873
Db 830 AYRAILHIPTRIOGLERALL 850

RESULT 9
US-08-022-835-2
; Sequence 2, Application US/08022835
; Patent No. 5420030
; GENERAL INFORMATION:
; APPLICANT: Reitz Jr., Marvin S.
; APPLICANT: Franchini, Genoveffa
; APPLICANT: Markham, Phillip D.
; APPLICANT: Gallo, Robert C.
; APPLICANT: Lori, Franco C.
; APPLICANT: Popovic, Mikulas
; APPLICANT: Gartner, Suzanne
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1 AND USES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: Eleventh Floor, 1615 L. Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/022.835
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/599,491
 FILING DATE: 17-OCT-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Scott, Watson T.
 REGISTRATION NUMBER: 26,581
 REFERENCE/DOCKET NUMBER: WTS/5683/78507/SRL
 TELEPHONE: (202) 861-3000
 TELEFAX: (202) 822-0944
 TELEX: 6714627 CUSH
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 856 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-022-835-2

Query Match 47.8%; Score 2211.5; DB 1; Length 856;
 Best Local Similarity 49.4%; Pred. No. 2.9e-173;
 Matches 440; Conservative 159; Mismatches 236; Indels 55; Gaps 23;

QY 5 MRAMG-KRNRK--LGILYIVMALIIPCLSSOLYATVYAGVPVWEDAPVLFCSADANLT 61
 DB 1 MRVKGIRRYQHWGWTGMLLGLMICSATEKLWTVYGVVWKEATVTLFCASDAKAY 60

QY 62 STEKHNVASQACVPDPHPHYLLTVNVDNFNIWENTAVQOMDIIISLWDQSLKPCIQ 121
 DB 61 DTEVHNVAQACVPDPNPQVQVWVNVTEFNEMKNMNMVQMHEDIIISLWDQSLKPCVK 120

QY 122 MTFMCIQMCTDIKN--NNTSGTENRTSSSE-----NPMKTCERNITVLKDKKQKAL 174
 DB 121 LPLCLVTLNCTDLRTNTNNTNANNNSSEGTIKGEMKNSFNITTSIRDKMQKEVAL 180

QY 175 FVVSDLTKLADNNTNTMYTLNCTNNTIKQACPKVSEPIPIYCAPAGVAFKCHSAE 234
 DB 181 LKLVLSI-DNDSTS--YRLISCNSTVITQACPKISEPIPIYCAPAGVAFKCHNDKK 237

QY 235 FNGTCKSNISVVTCTHGIKTVSTOLLNGLTSLKEIRIMKNSDSGKNIIVTLSSDI 294
 DB 238 FSGKSCANVSTVQCTHGIKTVSTOLLNGLTSLKEIRIMKNSDSGKNIIVTLSSDI 297

QY 295 EITCVRPGNNTVQEMKIGP--MAWYSM--ALGTGNSRSRVAYCOYNTTEKALKNTAER 351
 DB 298 QINCTRPYNNKR-KRIHIGPGRAEYTKNIQT-----IROAHCNISRAKWNVDLQIVSK 352

QY 352 YLELINTGNTMIFNRSQDSGVDEVTHLHFNCHGEFFYCNTESEFNTFLCNGTNC-N 410
 DB 353 LKEQFK-----NKTIVFQOS--SGDPEIVMHSFNCGGSEFFYCNTESEFNTFLCNGTNC-N 405

QY 411 NTQINSANGMIPCKLQVYRSMRGSGLYAPPICNLTCSHIGMILO-----MDAPW 466
 DB 406 NTTGSNN-NITLQCKIKIINMQEVGKAMYAPPIEQIKCSNITGLLTLRDGGKDTDT 464

QY 467 NKTENTFRPGDKMDIWRNELFKYKVVYKVPFSVAPTPIAREVIGTGHREKRAVGLGM 526
 DB 465 NDE-IFRPGGDMRDNRSSELYKYVVTIEPLGVAPTAKRRV---QREKRA-ATGA 518

QY 527 LFLGLVSAAGTGAATALTQVTHSVKIGVQOONLIRATQAOQLRLSVWGIQRL 586
 DB 519 LFLGLVSAAGTGAASVTLTQVQARLLSGVQOONLIRAEAOQHMQLQVWGIQQL 578

QY 587 ARLLALETIQOOLLNMGCKRLCYTSVKKNWETWNTTINQINWGLNLTWQEDQID 646
 DB 587 ARLLALETIQOOLLNMGCKRLCYTSVKKNWETWNTTINQINWGLNLTWQEDQID 646

DB 579 ARVLAVERLYKDOQLLGFWCGSKLICITTTVPWNASWN-KSLDDINNNWMTWQWERID 637
 QY 647 NVSSYIEEIOKAQVQOQONKLEDEWASLNWLDITKWLWYIKIATIIIVGALIGVR 706
 DB 638 NVTSIYSLLEKSTQOQNEQELLELDKASLNWFDITNWLWYIKIFIMIVGLVGLR 697

QY 707 IVMIVNLVRNIROGYQPLSLQIPTROQSEAEPTGRTGEGGDEGRP---RLIPSPQGF 763
 DB 698 IVFAVLIVNVROGYSPLSLQTPRPVPRGDPREGIEEGGERDRTSRLV---HGFL 754

QY 764 PLLATDLRTIILWSYHLLSNLISCTQVIVSHLRGLGLWILGOKIIDACRICAAYHVLQ 823
 DB 755 ALIWDLSLFLSYHHRDLTLTAARIV-----ELGRRGWEVLKYWNLLQYWSQE 806

QY 824 LKSAVSLNATAVAEGTDRVIEWLQRAGRAILHIPTRIQGLEALL 856
 DB 807 LKSAVSLNATAVAEGTDRVIEWLQRAGRAILHIPTRIQGLEALL 856

RESULT 10
 US-08-388-809-2
 ; Sequence 2, Application US/08388809
 ; Patent No. 5576000
 ; GENERAL INFORMATION:
 ; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,
 ; APPLICANT: GENOVEFPA; MARKHAM, PHILIP D. GALLO, ROBERT
 ; APPLICANT: C.; IORI, FRANCO C.; POPOVIC, MIKULAS; AND
 ; APPLICANT: GARTNER, SUZANNE
 ; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK, 3.5"
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/388,809
 ; FILING DATE: 15-FEB-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/022,835
 ; FILING DATE: 25-FEB-1993
 ; APPLICATION NUMBER: US 07/599,491
 ; FILING DATE: 17-OCT-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: LESLIE A. SERUNIAN
 ; REGISTRATION NUMBER: 35,353
 ; REFERENCE/DOCKET NUMBER: 2026-4092US2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 856 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-388-809-2

Query Match 47.8%; Score 2211.5; DB 1; Length 856;
 Best Local Similarity 49.4%; Pred. No. 2.9e-173;
 Matches 440; Conservative 159; Mismatches 236; Indels 55; Gaps 23;

QY 5 MRAMG-KRNRK--LGILYIVMALIIPCLSSOLYATVYAGVPVWEDAPVLFCSADANLT 61

Db 1 MRVGIIRNYOHWWGWTMLGLLMICSAEKLVTVYGVVWKEATITLFCASDAKAY 60
 QY 62 STEKHNVWASQACVPTDPTPEHYLLTNVDNFNIWENYVYQMOEDIIISLDQSLKPCIQ 121
 Db 61 DTEVHNWATQACVPTDPTPEHYLLTNVDNFNIWENYVYQMOEDIIISLDQSLKPCVK 120
 QY 122 MTFWCIOMNCTDIKN--NNTSGTENRTSSSE-----NPMKTCFENITTVLKDKKKEKQAL 174
 Db 121 LTPCLVTLNCTDURNTNTNNSANNSEGGIKGEMKNCNFNTTIRDKMQKEYAL 180
 QY 175 FVYSDTLKADNNNTMTYTLNCTNTTKQACPKVSFEPIIYICAPAGYAFKCNASAE 234
 Db 181 LYKLDIVSI--DNDS--YRLISCNVTQACPKISFEPIIYICAPAGYAFKCNKDK 237
 QY 235 FNGTGKSNISVYTCGHGKPTVSTQILNGLTSLKSKIRIMGNKNSDGSNGNIIVTLSSDI 294
 Db 238 FSGKSGCKNVSTVQCTHGRIPVSTQILLNGSLAEVEVIRSENFDTNAKTIIVHLNESV 297
 QY 295 EITCVRPGNNOTVOEMKIGP--MAYISM--ALGTGNSRSRVAYCOYNTTEWEKALKNTAER 351
 Db 298 QINCTRPYNKR--KRIHIGPGRAFYITKNIGT----IROAHCNISRAKWNWDTLRQIVSK 352
 QY 352 YLELINTEGNTMIFNRSQDSDVEVTHLHFNCHEFFCYNTSEMFNTFLCNGTNC-N 410
 Db 353 LKEQFK----NKTIVFNQS--SGGDPFVMSFNCGGEFFCYNTSPLFNSTW--NGNNTWN 405
 QY 411 NTQINSANGMIPCKLKQVVRSMRGSGLYAPPNGLNCTICSHITGMILQ----MDAPW 466
 Db 406 NTTGSNN-NITLOCKIKQIINNWEVEGKAMYAPPIEGQIRCSNITGLLITRDGKDTDT 464
 QY 467 NKTENTPRIGDMDKIDRNELFKYVVRKVPFSAVPTPIARPIVIGTTHREKRAVLGM 526
 Db 465 NDTE-IFRPGGDMRDNWRSLEYKVKVVTIEPLGVATPKAKRRV---QREKRA-AIGA 518
 QY 527 LFLGVLGAAGTGAATATLVQTHSVKIGVQOQDNLRAIQAOQELLSRWIGRQLR 586
 Db 519 LFLGVLGAAGTGAATATLVQTHSVKIGVQOQDNLRAIQAOQELLSRWIGRQLR 578
 QY 587 ARLLALETIONQQLNLNGLCKGRICVTSVKWNETWRNTNINQNGNTWQEMDOID 646
 Db 579 ARVLAVERLYKDDQLLGFWSGGLICTITVPVWASNN--KSLDDIWNNTWQWQEREID 637
 QY 647 NVSSTIIEIOKAQVQEQNEKLELDEWASLWNLNLDITKWLVIKIAIIVGALIGVR 706
 Db 638 NYSLSIYSLKESQTEKNEQELLELDKASLWNLNLDITKWLVIKIAIIVGALIGVR 697
 QY 707 IVMIVLNRNIRGOYPLSLQIPTRQOSEAETPGRTGEGGDEGRP---RLIPSPQGL 763
 Db 698 IVFVAVLSVNRVROGYSPLSLQIPTRPVPGRDPEGIEEGGERDRDTSGLV---HGFL 754
 QY 764 PLLYDRLTILWYSHLLSNLSGTQVISHRLGLWLGKIIDACRICAAVHYWLOE 823
 Db 755 ALLWDLRSLFYSYHRDLILLIARIV-----ELGRRGWELVKYWNLLQYWSQE 806
 QY 824 LQKATSILDTFAVANWTDIILGIQRLGRGILNTPRRVQCFERSLL 873
 Db 807 LKSSAVSLNATAVAEGTDRVLEVLQAGRAILHPIRQGLERAIL 856

RESULT 11
 US-08-647-714-2

; Sequence 2, Application us/08647714
 ; Patent No. 5869313
 ; GENERAL INFORMATION:
 ; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,
 ; APPLICANT: GENOVEFA; MARKHAM, PHILIP D. GALLO, ROBERT
 ; APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
 ; APPLICANT: GARTNER, SUZANNE
 ; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK, 3.5"
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/647,714
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/388,809
 ; FILING DATE: 15-FEB-1995
 ; APPLICATION NUMBER: US 08/022,835
 ; FILING DATE: 23-FEB-1993
 ; APPLICATION NUMBER: US 07/599,491
 ; FILING DATE: 17-OCT-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: LESLIE A. SERUNIAN
 ; REGISTRATION NUMBER: 35,353
 ; REFERENCE/DOCKET NUMBER: 2026-40920S2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 856 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-647-714-2

Query Match 47.8%; Score 2211.5; DB 2: Length 856;
 Best Local Similarity 49.4%; Pred. No. 2.9e-173;
 Matches 440; Conservative 159; Mismatches 236; Indels 55; Gaps 23;

QY 5 MRAMG-KRNRK--LGILYIVMALIIFCLSSQLYAVYAGVYVWEDAAAPVLCASDANLT 61
 Db 1 MRVGIIRNYOHWWGWTMLGLLMICSAEKLVTVYGVVWKEATITLFCASDAKAY 60
 QY 62 STEKHNVWASQACVPTDPTPEHYLLTNVDNFNIWENYVYQMOEDIIISLDQSLKPCIQ 121
 Db 61 DTEVHNWATQACVPTDPTPEHYLLTNVDNFNIWENYVYQMOEDIIISLDQSLKPCVK 120
 QY 122 MTFWCIOMNCTDIKN--NNTSGTENRTSSSE-----NPMKTCFENITTVLKDKKKEKQAL 174
 Db 121 LTPCLVTLNCTDURNTNTNNSANNSEGGIKGEMKNCNFNTTIRDKMQKEYAL 180
 QY 175 FVYSDTLKADNNNTMTYTLNCTNTTKQACPKVSFEPIIYICAPAGYAFKCNASAE 234
 Db 181 LYKLDIVSI--DNDS--YRLISCNVTQACPKISFEPIIYICAPAGYAFKCNKDK 237
 QY 235 FNGTGKSNISVYTCGHGKPTVSTQILNGLTSLKSKIRIMGNKNSDGSNGNIIVTLSSDI 294
 Db 238 FSGKSGCKNVSTVQCTHGRIPVSTQILLNGSLAEVEVIRSENFDTNAKTIIVHLNESV 297
 QY 295 EITCVRPGNNOTVOEMKIGP--MAYISM--ALGTGNSRSRVAYCOYNTTEWEKALKNTAER 351
 Db 298 QINCTRPYNKR--KRIHIGPGRAFYITKNIGT----IROAHCNISRAKWNWDTLRQIVSK 352
 QY 352 YLELINTEGNTMIFNRSQDSDVEVTHLHFNCHEFFCYNTSEMFNTFLCNGTNC-N 410
 Db 353 LKEQFK----NKTIVFNQS--SGGDPFVMSFNCGGEFFCYNTSPLFNSTW--NGNNTWN 405
 QY 411 NTQINSANGMIPCKLKQVVRSMRGSGLYAPPNGLNCTICSHITGMILQ----MDAPW 466
 Db 406 NTTGSNN-NITLOCKIKQIINNWEVEGKAMYAPPIEGQIRCSNITGLLITRDGKDTDT 464


```

QY 467 NKTENTFRPIGDMKDIWRNELEFKYKVVVAPFVAPTPVIGTGTTHREKRAVGLGM 526
Db 465 NDTE-IFRGGGDRNDWSELYKVKVWVIEPLGVAPTAKRRVV----QREKRA-AIGA 518
QY 527 LFGVLSAAGSTMGAAATALTQTHSVIKGIVQOQDNLRLAIOAQOELLRLSVGIRQLR 586
Db 519 LFLGFLGAAGSTMGAASTVLTQVQALLSGIVQOQNNLRLAIEAQOQLQVWGIKQLQ 578
QY 587 ARLLALETIONQOQLNLCGCKGRICYSVKWNETWNTTINQINWNTTWOEWQOQID 646
Db 579 ANVLAVERTKQOQLGFWGSGKLCITTVFVWNASWN-KSLDDIWNNTWQWEREID 637
QY 647 NVSYTIEIQKAOVQOQNEKKELEDEWASLWNLWLDITKWLVIKIAIIVGALIGVR 706
Db 638 NYSLSIYSLKESQTOQEKNEQELLELDKWSLWNNWDTNWLWIKIFIMVIGGLVLR 697
QY 707 IVMIVNLVNRIOGYQPLSQIPTROQSEATPGRTEGGGDEGRP----RLIPSPQGL 763
Db 698 IVFVLSVNRVROGYSPLSLQTPRPVPRGDRPEGIEEGGERDRDTSGLV----HGFL 754
QY 764 PLLYTDLRTIILWSYHLLSNLSIGTQVISHRLGLWILGOKIIDACRICAIVHWWQ 823
Db 755 AIIWDLRSFLFSYHRDLIAARIV------ELGRRGWELKYWNLLQYWSOE 806
QY 824 LOKSATSILDTFAVANWTDITLILGRLGRGILNIPRRVQGFERSLL 873
Db 807 LKSAVSLNATAVAAGTDRVIEVQLRAGRAILHIPRIQGLERALL 856

```

RESULT 12

```

us-08-463-210-11
Sequence 11, Application US/08463210
Patent No. 6001977
GENERAL INFORMATION:
APPLICANT: CHANG, Nancy T.
APPLICANT: GALLO, Robert C.
APPLICANT: WONG-STAAI, Flossie
TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan, L.L.P.
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
* COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,210
FILING DATE: 05-JUN-1995
CLASSIFICATION: 436
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 06/693,866
FILING DATE: 23-JAN-1985
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 06/659,339
FILING DATE: 10-OCT-1984
ATTORNEY/AGENT INFORMATION:
NAME: Serunian, Leslie A.
REGISTRATION NUMBER: 35,353
REFERENCE/DOCKET NUMBER: 2026-4193US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 863 amino acids
TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..863
; OTHER INFORMATION: /note= "env protein of HTLV-III"
US-08-463-210-11

```

```

Query Match 47.8%; Score 2211.5; DB 3; Length 863;
Best Local Similarity 49.3%; Pred. No. 3e-173;
Matches 427; Conservative 164; Mismatches 238; Indels 33; Gaps 16;
QY 20 IVMAIIICLSSQLXATVYAGVPVWEDAAPVLCFASDANLSTKHNWVWASQACVPTDP 79
Db 27 MLLGLMICSAEKLVVYVYGVVWKEATTLFCASDAKAYDEVHNVWATHACVPTDP 86
QY 80 TPHEYLITVNTDFNFWNVMEQOEDIISLWDSLAPCIOMFCIOMNCTDKNN-N 138
Db 87 NPOEVVLNVTFNFMKNDVMEQHEDIISLWDSLAPCVKLTPLCVSLACTLKNKTN 146
QY 139 TSGTENRTSSSNPMKTCENITTVLKDKKKEKQALFYVSDLTKLADNNTNTMTLNC 198
Db 147 TNSSSGRMIMEKEIKNCFSNISTIRKVQKEYAFFYKLDIPI-DNDTTS--YTLTSC 203
QY 199 NSTTIQACPKVSPPIPIYICAPAGYAFKCSNAEFNGTCKSNISVVTCTHGIKPTVS 258
Db 204 NTSVITQACPKVSPPIPIYICAPAGYAFKCSNAEFNGTCKSNISVVTCTHGIKPTVS 263
QY 259 TOLLNGTSLSEKIRIMGNISDSGKNIIVTLSSDIETCVPCNNQTVQBMKI--GPMA 316
Db 264 TOLLNGSLAEVIVRSANETDNKIIIVQLOSVEINCTRP--NNTRKSIIRIQGPR 322
QY 317 WYSMALGTGSNRSVAYCOYNTTEWEKALKNTAERYLELINTEGNTMIFNSQDSDV 376
Db 323 AF-VTIGKIGN-MRQAHCHNISRAKWNNTLQIDSKLREQFN--NKTIIFKQSSGDPE 377
QY 377 EVTHLFNCHGEFFYCYNTSEMFNYTELNGT----NCNNTQINSANGMIPCKLKQVRS 432
Db 378 IVTH-SFNGCGEYCYNSTOLFSTWF-NSTWTKGNSNTEGSDTIT--LPCRIKOINN 433
QY 433 WMRGSGLYAPPPIGNLTCSHITGMILQMDAPWNKTEN-TFRPIGDMKDIWRNELEFKY 491
Db 434 WQEVGKAYAPPISGOIRCSNITGLLLTRDGNSSNESEIFRPGGDMRDNWRSLEYKY 493
QY 492 KVRVKPFSVAPTPIARPVIGTGTTHREKRAVGLMLFLGVLSAAGSTMGAAATALTQ 551
Db 494 KVKIEPLGVAPTAKRRVV----QREKRAVIGALFLGLGAAGSTMGAAATALTQ 549
QY 552 SVKGVQOQDNLRLAIOAQOELLRLSVGIRQLRLALETIONQOQLNLCGCKGR 611
Db 550 QLLSGIVQOQNNLRLAIEAQOQLQVWGIKQLQVWGIKQLQVWGIKQLQVWGIK 609
QY 612 ICYTSVKWNETWNTTINQINWNTTWOEWQOQIDINVESTIYEETQKAOVQOQNEK 671
Db 610 ICTATVPWNASWN-KSEQIWNNTWMEWDREINNYTSLIHSIEESQOQNEKQELL 668
QY 672 ELDEWASLWNLWLDITKWLVIKIAIIVGALIGVRIVMLVNLVNRIOGYQPLSLQ 731
Db 669 ELDKWASLWNLWNTNWLWYIKLFIMVIGGLVGLRIVFVLSVNRVQGSPLSFOT 728
QY 732 RQSEAEETPRTGEGGDEGRPLIPSPQGLPLDLYTDLRTIILWSYHLLSNLSIG 791
Db 729 PTPRGDPEEGIEEGGERDRSRILVNGSLALIWDDLRLSLCLFSYHRLDLLIV 788
QY 792 ISHLRLGLWILGQKIIDACRICAIVHWWQELQKSLDITFAVANWTDITLILGI 851
Db 789 VE-----LLGRGWELKYWNLLQYWSLKNASVSLNATAVAAGTDRVIEVQ 841

```


APPLICATION NUMBER: US/08/388,809
 FILING DATE: 15-FEB-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/022,835
 FILING DATE: 25-FEB-1993
 APPLICATION NUMBER: US 07/599,491
 FILING DATE: 12-OCT-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: LESLIE A. SERUNIAN
 REGISTRATION NUMBER: 35,353
 REFERENCE/DOCKET NUMBER: 2026-4092US2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 751-6840
 TELEFAX: (212) 751-6849
 TELEX: 421792
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 857 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-388-809-4

Query Match 47.8%; Score 2211; DB 1; Length 857;
 Best Local Similarity 49.1%; Pred. No. 3.2e-173;
 Matches 437; Conservative 159; Mismatches 240; Indels 54; Gaps 22;

QY 5 MRAMG-KRNRK--LGLIYIVMALIIPCLSSSQLYATYAGVPVWEDAAPVLFCSANLIT 61
 DB 1 MRVKGIRRYQHGWGTLGLMISATEKLMVTVYGVVPVKEATITLFCASDAKAY 60

QY 62 STEKHNVASQCVPTDTPHEYLITNTVNFNIWENYEQMEDIIISLWQSLKPCIQ 121
 DB 61 DTEVHNWATHACTPTDFNPOEVELNVNTEFNWNNKNNVQEHEDIISLWQSLKPCVK 120

QY 122 MTFMCIONMCTDIKN--NNTSGTENRTSSE-----NPMKTCFNTIYVLDKKEKQAL 174
 DB 121 LTPLCVTLNCTDLNTNTNNTANNNSSEGTIKGKNGKNSFNITTSIRDKMQKEVAL 180

QY 175 FVSDLTKLADNNTNTMYTLINCNTTIKQACPKVSEPIPIYCAPAGYAFKCSAE 234
 DB 181 LYKLDIVSI---NNDSTYSLSCNTSVITQACPISEPIPIHVCAPAGFAILKNDKK 237

QY 235 FNGTGCNSIVVTCFHGKIPVSTQILINGPLSKHTRIMGKNISDGRNIIVTSSDI 294
 DB 238 FSGKSGCKNVSTVQCTHGRVPSVQQLLNGSLAEVEVIRSENFNDNAKTIIVHLSNV 297

QY 295 EITCVRPNGNQTVOEMKIGP-WANTSM--ALGTGNSRNVAYCOYNTTEWEKALKNTAER 351
 DB 298 QINCTEPNKR-KRIHIGPGRAFYTKNIIGT----IRQAHCNISRAKNWDLRQIVSK 352

QY 352 YLELINTEGNTMIFNRSODGSQVETHLHFNCHGEFFYCNTESEMPNTYFLNGTNC-N 410
 DB 353 LKEQFK----NKTIVFNGS--SGDPEIVMSHFGCGGEFFYCNTESEMPNTYFLNGTNC-N 410

QY 411 NQTOSINSANGMLPCKLVYRSMRGSGLYAPPFPGNLTICSHITGMLQ---MDAPW 466
 DB 406 NITGSNN-NITLQCKIKGIINMOEYKATYAPPIEGQIRGSSNITGLLIRFDGKDTDT 464

QY 467 NKTENYRPTGGDKMDIWRNELFKYKVRVPSPVAPTPIARPVIGTGTTHREKRAVLGM 526
 DB 465 NDTE-IFRPGGDMRDNWSELYKYKVVIEPLGVAPTAKARVY---QREKRA-AIGA 518

QY 527 LFLGLVSAAGSTGAATATLVQTHSVIGIVQOQDNLRAIQAOQLRLSVNGIRQLR 586
 DB 519 LFLGLVSAAGSTGAASVTLTVQARLLUSGIVQOQDNLRAIQAOQLRLSVNGIRQLR 578

QY 587 ARLLALETLIONOOLLNMGCKRGLICTSVKWNTEWNTNINQINWNTWQEWQOQID 646
 DB 579 ARILAVETLKQQLLGGWCGSKLICITTPWNASRN-KSLDDIWNNTWQWQEREID 637

QY 647 NVSSTIYEIQAAQVQEQNECKKLELDWASLNNWLDITKWLWYIKAIIVGALIGYR 706

DB 638 NYTSLIYSLEKSTQTOENNEQELLELDKWSLWNNWFDITNWLWYIKIFIMVGLVGLR 697
 QY 707 IVMLVNLVNRNQYQVPLSLQIPTRQOSEATPGTGEQGGDEGRP---RLIPSPQGF 763
 DB 698 IVFAVLISVNRVQYSPSLQTRPPVPRGDRPEEGIEEGEDRDTSRLV---HGFL 754
 QY 764 PLYLTDRTIILWSYHLLSNLISGTQTVSHRLGLMLGKILQKIDACRICAAYIHWLQ 823
 DB 755 AIIWDLRSLEFLFSYHLLRDLILLTAARIVE-----LLGRRGWEVLKYWNLLQYWSQE 807
 QY 824 LOKSATSLIDTFAVANWNTDDIILGQIRLGRGILNIPRRVROGERSLL 873
 DB 808 LASSAVSLNATDIAVAEGTDRVIEVQLQRAIRAILHPTIIRQGLERALL 857

RESULT 15
 US-08-647-714-4
 ; Sequence 4, Application US/08647714
 ; Patent No. 5869313
 ; GENERAL INFORMATION:
 ; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,
 ; APPLICANT: GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT
 ; APPLICANT: G.; IORI, FRANCO C.; POPOVIC, MIKULAS; AND
 ; APPLICANT: GARTNER, SUZANNE
 ; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK, 3.5"
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/647,714
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/388,809
 ; FILING DATE: 15-FEB-1995
 ; APPLICATION NUMBER: US 08/022,835
 ; FILING DATE: 25-FEB-1993
 ; APPLICATION NUMBER: US 07/599,491
 ; FILING DATE: 17-OCT-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: LESLIE A. SERUNIAN
 ; REGISTRATION NUMBER: 35,353
 ; REFERENCE/DOCKET NUMBER: 2026-4092US2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 751-6840
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 857 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-647-714-4

Query Match 47.8%; Score 2211; DB 2; Length 857;
 Best Local Similarity 49.1%; Pred. No. 3.2e-173;
 Matches 437; Conservative 159; Mismatches 240; Indels 54; Gaps 22;

QY 5 MRAMG-KRNRK--LGLIYIVMALIIPCLSSSQLYATYAGVPVWEDAAPVLFCSANLIT 61
 DB 1 MRVKGIRRYQHGWGTLGLMISATEKLMVTVYGVVPVKEATITLFCASDAKAY 60

Db 1 MRVKGIRRNQHWGNGTMLGLLMLCSATEKWLWTVYVGVVWVWKEATTTLFCASDAKAY 60
 QY 62 STEKHNVWASQACVPTDPDPHEYLNTDNTFNWENYVWQEDIIISLWDSLAPCIQ 121
 Db 61 DTEVHNWATHACVPTDPDPQVELVNTFNWNNVWQEHEDIISLWDSLAPCVK 120
 QY 122 MTFMCOMNCTDIKN--NNTSGTENTSSSE-----NPMKTCFNTTIVLKDKKKEKQAL 174
 Db 121 LTPCLVTLNCTDLNNTNTNNSANNNSSEGTIKGEMKNCSEFNITTSIRDKMREYAL 180
 QY 175 FYVSDTLTKLADNNTMTYLINCNTTIKQACPKVSFPIPIYCAPAGYAIFKCNASAE 234
 Db 181 LYKLDVSI---NNDSTYSYLSNCTSVITQACPKISFEPIPIHYCAPAGFAILKCNCKK 237
 QY 235 FNGTGKSNISVVTCHGKIPVTSTQILNGLSKIRIMGKNISDSGKNIIVTLSSDI 294
 Db 238 FSGKSGCKNVSTVQCHGIRPVYSTQLLNGSLAEVIRSENFNDNAKTIIVHLNESV 297
 QY 295 EITCVRPNGNQVOEMKIGP--MAWYSM--ALGTGSNRSRVAYCOYNTTEWEKALKNTAER 351
 Db 298 QINCTRPYNKR-KRTHIGPGRAYTTKNIIGT---IRQAHCNISRAKWNQDLRQIVSK 352
 QY 352 YLELINNTEGNTMIENRSQDSDVEVTHLFNCHGEFFCYTSEMPTYFLCNGTNC-N 410
 Db 353 LKEQFK---NKTIVFNQS--SGGDPFVHMSFNGGGEFFCYTSPFNSTW--NGNNTWN 405
 QY 411 NTQSIANSANGMIPCKLKQVYVSRWVGSGLYAPPPIGNLTCISIHITGMILQ-----MDAPW 466
 Db 406 NTGGSNN-NITLQCKIKQIINMWQEVGKAIYAPPIEGQIRCSNITGLLLTRDGGKDDT 464
 QY 467 NKTENTFRPIGGDMKDIWRNELFKYKVVVPKFSVAPTPIAREVIGTGTHREKRAYGLGM 526
 Db 465 NDTE-IFRPGGDMRDNRSSELYKYKVTIEPLGVAPTRAKRRV---QREKRA-AIGA 518
 QY 527 LFLGVLSAAGSTGAATALTQVTHSVIKGIVQOQDNLLRAIOAQOQLLSLWSWGIQRLR 586
 Db 519 LFLGFLGAGSTGAASVTLTVQARLLSGIVQOQNNLLRAIEAQOQMLQLTVWGIKQLQ 578
 QY 587 ARLEALETLLIONQLNLWCKGRLCYTSVKWNETWNTNINQIWGNLTWOEWDOQID 646
 Db 579 ARILAVERYLKDQOLLGWCCKLICTTVPWNASWSN-KSLDDIWNNTWQWQEREID 637
 QY 647 NVSTIYEETQKAQVOQOQNEKKLELDEWASLWNLWDITKWLWYIKIAIIVGALIGVR 706
 Db 638 NYTSLIYSLEKSQTCQEMNEQELLEDKWASLWNFDTIINWLMYIKIFIMIVGLVGLR 697
 QY 707 IVMIVNLVRNIRGQYQPLSLQIPTQQOSEAFPGTGECCGDEGRP---RLIPSPQGL 763
 Db 698 IVFAVLISIVNRVQGYSPLSLQTPRPVPRGPRGEGIEEGGERDRDTSGRLV---HGFL 754
 QY 764 PLYTDLRTIILWSYHLLSNLSGTQTVISHRLGLWILGOKIIDACRICAAVIHWLOE 823
 Db 755 AITWDLRSLEFSYHRLDLILLIARIVE-----LIGRRGEVFLKYWNWLLQYWSQE 807
 QY 824 LQKSATSLIDTFVAVANWTDIILGIQIRGLINIPRVRQGFERSLL 873
 Db 808 LKSSAVSLLNATDIAVAEGTRVIEVQLRAGRAILHIPRIROGLERALL 857

Search completed: January 25, 2001, 11:00:43
 Job time: 559 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2001, 10:51:24 ; Search time 43.17 Seconds
(without alignments)
2370.223 Million cell updates/sec

Title: US-08-911-824-61
Perfect score: 4622
Sequence: 1 MIVTPRAMGRKNRKLGIYI.....GRGILNPRVRQGFERSLL 873

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_prodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3859	83.5	879	12 Q9WU9	Q9WU9 human immun
2	3670.5	79.4	876	12 Q9YKB2	Q9YKB2 human immun
3	3664.5	79.3	872	12 O57072	O57072 human immun
4	3664	79.3	871	12 O57073	O57073 human immun
5	3654.5	79.1	900	12 Q9QN28	Q9QN28 human immun
6	3650	79.0	871	12 O57074	O57074 human immun
7	3632	78.6	863	12 Q77377	Q77377 human immun
8	3577.5	77.4	860	12 Q93091	Q93091 human immun
9	3414	73.9	877	12 Q76576	Q76576 human immun
10	3381.5	73.2	876	12 Q79670	Q79670 human immun
11	2383	51.6	840	12 Q9QNX7	Q9QNX7 human immun
12	2370	51.3	863	12 O40360	O40360 human immun
13	2356.5	51.0	860	12 Q9QRW2	Q9QRW2 human immun
14	2356	51.0	862	12 Q9WLG8	Q9WLG8 human immun
15	2342.5	50.7	859	12 Q9WJ11	Q9WJ11 human immun
16	2338.5	50.6	537	12 Q9IED8	Q9IED8 human immun
17	2334.5	50.5	861	12 Q73366	Q73366 human immun
18	2334	50.5	859	12 Q76128	Q76128 human immun
19	2332.5	50.5	864	12 Q70210	Q70210 human immun

20	2330	50.4	870	12 Q9YMY6	Q9YMY6 human immun
21	2328.5	50.4	536	12 Q9IEE5	Q9IEE5 human immun
22	2328.5	50.4	845	12 Q9ID89	Q9ID89 human immun
23	2324.5	50.3	855	12 Q9YUV3	Q9YUV3 human immun
24	2324.5	50.3	861	12 Q73360	Q73360 human immun
25	2323.5	50.3	855	12 Q70016	Q70016 human immun
26	2321.5	50.2	851	12 Q73365	Q73365 human immun
27	2319.5	50.2	853	12 Q70202	Q70202 human immun
28	2317.5	50.1	856	12 Q9WLG6	Q9WLG6 human immun
29	2317	50.1	847	12 Q9WIS1	Q9WIS1 human immun
30	2316	50.1	854	12 Q97016	Q97016 human immun
31	2316	50.1	857	12 P90249	P90249 human immun
32	2315.5	50.1	868	12 Q80225	Q80225 human immun
33	2314.5	50.1	853	12 Q9WBT5	Q9WBT5 human immun
34	2314.5	50.1	857	12 Q97065	Q97065 human immun
35	2314.5	50.1	858	12 O11945	O11945 human immun
36	2314	50.1	854	12 Q72744	Q72744 human immun
37	2312.5	50.0	851	12 Q73361	Q73361 human immun
38	2312.5	50.0	854	12 Q9YV14	Q9YV14 human immun
39	2312	50.0	857	12 Q9YU2	Q9YU2 human immun
40	2310	50.0	861	12 Q9WY9	Q9WY9 human immun
41	2308	49.9	840	12 Q9WBT0	Q9WBT0 human immun
42	2308	49.9	846	12 Q9ID98	Q9ID98 human immun
43	2306.5	49.9	849	12 Q9YU11	Q9YU11 human immun
44	2305	49.9	854	12 Q9WSE4	Q9WSE4 human immun
45	2304.5	49.9	854	12 Q9WC85	Q9WC85 human immun

ALIGNMENTS

RESULT	1
Q9WU9	
ID	Q9WU9 PRELIMINARY; PRT; 879 AA.
AC	Q9WU9;
DT	01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE	GP160 PRECURSOR.
GN	ENV.
OS	Human immunodeficiency virus type 1.
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX	NCBI_TaxID=11676;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-VI686;
RX	MEDLINE=99223950; PubMed=10207543;
RA	Janssens W., Heyndrickx L., Van der Auwera G., Nkengasong J.,
RA	Beirnaert E., Vereecken K., Coppens S., Willems B., Franssen K.,
RA	Peeters M., Ndumbe P., Delaporte E., Van Der Groen G.,
RT	"Interpatient genetic variability of HIV-1 group O.";
RL	AIDS 13:41-48(1999).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN-VI686;
RX	MEDLINE=96426454; PubMed=8828748;
RA	Delaporte E., Janssens W., Peeters M., Buve A., Dibaba G.,
RA	Perret J.L., Ditsambou V., Georges Courbot M.C., Georges A.,
RA	Bourgeois A., Samb B., Henzel D., Heyndrickx L., Franssen K.,
RA	Van Der Groen G., Larouz B., Mbe J.R.;
RT	"Epidemiological and molecular characteristics of HIV infection in
RL	Gabon, 1986-1994.";
RL	AIDS 10:903-910(-1996).
DR	EMBL; X96526; CAJ65373.1; -
DR	INTERPRO; IPR000287; -
DR	INTERPRO; IPR000777; -
DR	PFAM; PF00516; GP120; 1.
DR	PFAM; PF00517; GP41; 1.
KW	Signal.
FT	SIGNAL 1 29
FT	CHAIN 30 524
FT	CHAIN 132 156
FT	CHAIN 164 200
FT	POTENTIAL.
FT	GP120.
FT	V1.
FT	V2.

FT	CHAIN	302	335	V3.
FT	CHAIN	395	427	V4.
FT	CHAIN	458	475	V5.
FT	CHAIN	525	879	GP41.
SEQ	SEQUENCE	879 AA; 99099 MW; BC6422ADDF1A3409 CRC64;		

Query Match		83.5%;	Score 3859;	DB 12;	Length 879;
Best Local Similarity		82.5%;	Pred. No. 0;		
Matches 730;		Conservative 58;	Mismatches 79;	Indels 18;	Gaps
QY	1	MIVTMRANGKRNKLGILYIVMALIIPCLSSSOLYATYVYAGVPWEDAAPVLFASDANL	60		
DB	1	MIVTMRVNEKKKKLWILCMIALIIPCLSSNHLIATYVYAGVPWEDANPVLFCASDANL	60		
QY	61	TSTKEHNWASCAVPTDPTPHEVLLTNTVDFNFIWENYWEQOEDIISLWDSQSLKPCI	120		
DB	61	TSTQHNWASQACVPTDPTPLEYPLKNVDFNFIWKNYWEQOEDIISLWESQSLPCV	120		
QY	121	QMTFCIQMNCOTDKNNNTSGTNRNTS-----SSENPMKTCFNFNTVTLKDKKQQA	173		
DB	121	QMTFLCVMQNCNNTSNTSSINATSSDTSKDTSSNSVSKOCEFNTVTLKDKKQQA	180		
QY	174	LFYYSDLTKLADNNTNTMYTLINCNSITIKQACPVSFEPIPIYCAPAGYAFKNSA	233		
DB	181	LFYYSDLVK-TDN--SPMTYFLINCNSITIKQACPVSFEPIPIHYCAPAGYAFKCNNT	236		
QY	234	EFNGTGKCSNIVTCTHGIKPTVSTQILNGTLSKEKIRIMGNKISDSGNKIIVTLSSD	293		
DB	237	EFNGTGKCNIVTCTHGIKPTVSTQILNGTLSKEKIRIMGNKISDSGNKIIVTLNSS	296		
QY	294	IEITCVRPGNNQTVQEMKIGPMWYSMALGTGSNRSRVAYCQYNTTEWEKALKNTAERYL	353		
DB	297	ISMTCEPGRNH-TVOEMKIGPMWYSMGLEENKTSRRAYCRYNATDWEKALKQWTERYL	355		
QY	354	ELINNTGENT-TMIFNSQDQSDVEVTHLHFNCHGEFFYCYNTSEMPNTTFLCNGTNCNNT	412		
DB	356	ELVNTSNKTVTMIFNQSNDG-DPEVTHLHFNCHGEFFYCYNTSRMFNTSFCSCTSCNSS	414		
QY	413	QSINSANG--MIPCKLKQVRSWNRGSGLYAPPDPGNLTCSHITGTMILQMDAPWNKTE	470		
DB	415	QNNNATHNDTQIPCRLLKQVRSWNRGSGLYAPIRGNLTCSHITGTMILQMDAPWNKSN	474		
QY	471	NTFRPIGDMKDIWRNELFKYKVVVRVFPFSVAPTPIARPVIGTGTTHREKRAVGLMFLG	530		
DB	475	TTFRPVGGDMKDIWRTELFKYKVVVRVFPFSVAPTPIARPVIGTGTTHREKRAVGLMFLG	534		
QY	531	VLSAAGSTMGAATAATVQTHSVIKIGIVQOQDNLRLRAQAOELLRLSVGIRQIRARLL	590		
DB	535	VLSAAGSTMGAATAATVQAHTLIKIGIVQOQDNLRLRAQAOQLLRLSVGIRQIRARLL	594		
QY	591	ALETLIQNOQLNLWGCKGRILICTSVKWNTEWNTNTIN--QIWGNLTWQEWQOQIDNV	648		
DB	595	ALETLIQNOQLNLWGCKGRILICTSVQWNKTWTNKSVDLENWGNLTWQEWQOQIDNI	654		
QY	649	SSTIIEIQKAQVOOQNEKKLLEDEWASLWNWLDITKWLWYIKIAIIVGALIGVRIV	708		
DB	655	SSTIYDEIQKAQVOOQNEQKLLLEDEWASITWNWLDITKWLWYIKMAIIVGALVGRIV	714		
QY	709	MIVLNLVRNIRQYQPLSLQIPTRQOSEATPPGTGEGGDEGRPLIPSPQGFPLPLYT	768		
DB	715	MIIUNLVRNIRQYQPLSLQIPTRHQSEATPPGTGEGGDEGRPLIPSPQGFPLPLYT	774		
QY	769	DLRTIILWSYHLLSNLISGTOVTSIHLRGLWILGQKIIDAACRIACAIVHWLQELQKSA	828		
DB	775	DLRTIILWYHLLSNLASGIQKASHLGLGLWILGQKIISACRIACAIVTQWELQNSA	834		
QY	829	TSLDITFAVANWTDIIILGIQRLGRGILNIPRRVROGPFERSLL	873		
DB	835	TSLDITLAVTVANNWDGIIILGIQIRGIRNIPTRIRQGLERSLL	879		

RESULT 2

Q9YKB2	PRELIMINARY;	PET;	876 AA.
ID	Q9YKB2		
AC	Q9YKB2;		
DT	01-MAY-1999 (T-EMBLrel. 10, Created)		
DT	01-MAY-1999 (T-EMBLrel. 10, Last sequence update)		
DT	01-MAY-2000 (T-EMBLrel. 13, Last annotation update)		
DE	GP160 PRECURSOR.;		
GN	ENV.		
OS	Human immunodeficiency virus type 1.		
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
OX	NCBI_TaxID=11676;		
	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=CA9;		
RC	MEDLINE=9923950; PubMed=10207543;		
RX	Janssens W., Heyndrickx L., Van der Auwera G., Nkengasong J.,		
RA	Beirnaert E., Vereecken K., Coppens S., Willems B., Fransen K.,		
RA	Peeters M., Ndumbe P., Delaporte E., Van Der Groen G.;		
RT	"Interpatient genetic variability of HIV-1 group O.;"		
RL	AIDS 13:41-48(1999).		
DR	EMBL; X96522; CAA65369.1; -		
DR	INTERPRO; IPR000328; -		
DR	INTERPRO; IPR000777; -		
DR	PFAM; PF00516; GP120; 1.		
DR	PFAM; PF00517; GP41; 1.		
KW	Signal.		
FT	SIGNAL	1	29
FT	CHAIN	30	524
FT	CHAIN	132	156
FT	CHAIN	158	199
FT	CHAIN	301	335
FT	CHAIN	394	425
FT	CHAIN	458	474
FT	CHAIN	525	876
SQ	SEQUENCE	876 AA;	998979 MW; B8E4D1B2EC656C9F CRC64;

Query Match	79.4%	Score	3670.5	DB 12	Length	876			
Best local Similarity	80.1%	pred. No.	1.5e-295						
Matches	707	Conservative	63	Mismatches	96	Indels	17	Gaps	12

QY	1	MIVTMRAGRRNRKLGILYIVMALIIPCLSSQLYATYIAGVPWDEADAPVLFCSADANL	60
DB	1	MTVTMKVMEKNRKLWLYIVMALIIPCLSSKQYATYIAGVPWGDADAPVLFCTSDVNL	60
QY	61	TSTEKHINWASQACVPDPPTPHEYLNTDNFNINWENYVEQMQEDIISLWDQSLKPCI	120
DB	61	TSTEKHINWASQACVPDPPTPHEYPLHNVDDFINWNNYVEQMQEDIISLWDQSSPCV	120
QY	121	QMTFMCIQMNC-TDIKNNNTSGTENRTSSSENPMKTCFENITVTKDKKKKQALFVSD	179
DB	121	QMTFLCQVMCTPQPQNSNTSSSENVTISEPNMKQCFENITVTKDKKKKQALFVSD	180
QY	180	LTKLADNNTTN---TWYTLINCNSTTIKQCPKVSFEPIPIYTCAPAGYAIFKCNSAEFN	236
DB	181	LVKL--NNTGNGNDMTYTLINCNSTTIKQCPKVSFEPIPIHYCAPAGYAIFKCNNTNFN	238
QY	237	GTGKCSNIYVTCGTHGKIPVSTQILIINGLTLSKEKIRIMKNTSDSGKNIIVTLSSDIEI	296
DB	239	GTGKCHNITVVTCHGKIPVSTHLLINGLTLSGKIRIMKNTNSAKNIIVTLNVTYDI	298
QY	297	TCVRPGNNOTVQEMKIGPMAWYSMALGTGS-NRSRVAYCQYNTEWEKALKNTAERYLEL	355
DB	299	TCERPGNH-TVQETIRIGPLAWYSMGIEKKNKSRLAYCNYNTDMKRALKQTAERYLEL	357
QY	356	INNTGENTMIFNRSQDGSQVETHLWFNCHGGEFFCYNTSEMENNYTFLCNGTCNNNTQSI	415
DB	358	INYT-NSVNMTFNHSIAGGDDETIMFNCHGGEFFYCNISKMFNYFECCNNTTCQG-NNI	415
QY	416	NSANG--MIPCKLQOVYRSMWRGSGSLYAPPINPGLNTCISHITGMILQMDAPWNKT-ENT	472
DB	416	TNDNGTDWICRLKOVYRSMWRGSGSLYAPPINRGTNITCRSNITGMHQLDRSNSSRDNT	475

DB 416 TNDNGTDWIPCRLRQVVRSMRGGGLYAPPIRGNTCRSNTTGMIMQLDRPWNSSRDNT 475


```
Query Match          79.3%; Score 3664; DB 12; Length 871;
Best Local Similarity 78.7%; Pred. No. 5e-295;
Matches 693; Conservative 72; Mismatches 98; Indels 18; Gaps 9;

QY 1 MIVTMRGKRNRKIGIYIVMALIIPCLSSOLYATYAGVPVWEDAAPVLFASDANL 60
DB 1 MIVTMRGKRNRKIGIYIVMALIIPCLSSOLYATYAGVPVWEDANPVLFASDANL 60
QY 61 TSTEKHNVASQACVPTDPTPHEYLNTVNTDNFNINWYMEQMEDIIISLWQSLKPCI 120
DB 61 TSTEKHNVASQACVPTDPTPHEYLNTVNTDNFNINWYMEQMEDIIISLWQSLKPCI 120
QY 121 QMTFCIOMNCTDIKNNTSGT-----ENRTSSSEPMKTCENIT 161
DB 121 QMTFCIOMNCTDIKNNTSGT-----ENRTSSSEPMKTCENIT 161
QY 181 TKLADNNTNTMYTLNCSNTHIKQACPKVSEPIPIYCAPAGYAIKCNASAEFNGTK 240
DB 181 TKLADNNTNTMYTLNCSNTHIKQACPKVSEPIPIYCAPAGYAIKCNASAEFNGTK 240
QY 241 CSNISVVTCTHGIRKPTVSTQILNGLTSKEKIRIMGNISDSGNKIIVTLSSDIEITCVR 300
DB 241 CSNISVVTCTHGIRKPTVSTQILNGLTSKEKIRIMGNISDSGNKIIVTLSSDIEITCVR 300
QY 301 PGNNTVQEMKIGPMANYSMAL-GTGSNRSRVAYCOYNTTEWEKALKNTAERYLELINT 359
DB 301 PGNNTVQEMKIGPMANYSMAL-GTGSNRSRVAYCOYNTTEWEKALKNTAERYLELINT 359
QY 360 EGNMTMIFNRSDGSDVEVTHLHFNCHGEFFYCNTSEMFNTFLCNGTNCN--NFQSNIS 417
DB 360 EGNMTMIFNRSDGSDVEVTHLHFNCHGEFFYCNTSEMFNTFLCNGTNCN--NFQSNIS 417
QY 418 ANGMIPCKLQVYRSMRGGSLYAPPIGNLTCTISHITGMILQMDAPWNTKTE-----NT 472
DB 418 ANGMIPCKLQVYRSMRGGSLYAPPIGNLTCTISHITGMILQMDAPWNTKTE-----NT 472
QY 473 FPIGDMKDIWRNELFKYVVRKVPFVAPTPPIARPVIGTTHREKRAVGLGMLFLGVL 532
DB 473 FPIGDMKDIWRNELFKYVVRKVPFVAPTPPIARPVIGTTHREKRAVGLGMLFLGVL 532
QY 533 SAAGSTMGAATAATYVQHSVIGVQOQDNLRAIOAQOELLRLSVGIRQLRARLLAL 592
DB 533 SAAGSTMGAATAATYVQHSVIGVQOQDNLRAIOAQOELLRLSVGIRQLRARLLAL 592
QY 593 ETILNQOQLNLNKGCKRLICHTSVKNWETWNTNINQINGNLFWQEDQIDNVSSTI 652
DB 593 ETILNQOQLNLNKGCKRLICHTSVKNWETWNTNINQINGNLFWQEDQIDNVSSTI 652
QY 653 YEEIOKAOVQOQNEKKLELDEWASLWNNLDTIKWLTIKIATIIYGALIGRVIMVL 712
DB 653 YEEIOKAOVQOQNEKKLELDEWASLWNNLDTIKWLTIKIATIIYGALIGRVIMVL 712
QY 713 NLVRNIRQYQPLSLQIPTRQOSEATPGTGGGDEGRPLIPSPQGFLLPYLTDLRT 772
DB 713 NLVRNIRQYQPLSLQIPTRQOSEATPGTGGGDEGRPLIPSPQGFLLPYLTDLRT 772
QY 773 IILWSYHLLSNLSSGQTQVISHRLGLNLGOKIIDACRICAAVYHWLQELQKATSIL 832
DB 773 IILWSYHLLSNLSSGQTQVISHRLGLNLGOKIIDACRICAAVYHWLQELQKATSIL 832
QY 833 DTFAVAVANWTDDIIGIQRGLRGIILNIPRRVROGFERSILL 873
DB 833 DTFAVAVANWTDDIIGIQRGLRGIILNIPRRVROGFERSILL 873
QY 873 DTFAVAVANWTDDIIGIQRGLRGIILNIPRRVROGFERSILL 873
DB 873 DTFAVAVANWTDDIIGIQRGLRGIILNIPRRVROGFERSILL 873

RESULT 5
ID Q9QN28 PRELIMINARY; PRT; 900 AA.
AC Q9QN28;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
```

01-JUN-2000 (TREMBLrel. 14, Last annotation update)

ENV. PROTEIN.

Human immunodeficiency virus type 1.

Viruses; Retroviruses; Retroviridae; Lentivirus.

NCBI_TaxID=11676;

[1]

SEQUENCE FROM N.A.

STRAIN=97CA-MP645M/O;

Peeters M., Liegeois F., Torimino N., Bourgeois A., Mpoudi E., Vergne L., Saman E., Delaporte E., Saragosti S.;

Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient;

J. Virol. 73:7368-7375(1999).

EMBL; AJ239083; CAB53242.1; -

INTERPRO: IPR000328; -

INTERPRO: IPR000777; -

PFAM: PF00516; GP120; 1.

PFAM: PF00517; GP41; 1.

SEQUENCE 900 AA; 100967 MW; FDB5E726A2537966 CRC64;

Query Match 79.1%; Score 3654.5; DB 12; Length 900;

Best Local Similarity 76.9%; Pred. No. 3.2e-294;

Matches 694; Conservative 72; Mismatches 104; Indels 33; Gaps 8;

QY 1 MIVTMRGKRNRKIGIYIVMALIIPCLSSOLYATYAGVPVWEDAAPVLFASDANL 60

DB 1 MIVTMRGKRNRKIGIYIVMALIIPCLSSOLYATYAGVPVWEDANPVLFASDANL 60

QY 61 TSTEKHNVASQACVPTDPTPHEYLNTVNTDNFNINWYMEQMEDIIISLWQSLKPCI 120

DB 61 TSTEKHNVASQACVPTDPTPHEYLNTVNTDNFNINWYMEQMEDIIISLWQSLKPCI 120

QY 121 QMTFCIOMNCTDIKNNTSGT-----ENRTSSSEPMKTCENIT 161

DB 121 QMTFCIOMNCTDIKNNTSGT-----ENRTSSSEPMKTCENIT 161

QY 162 TVLKDKEKKQALFYVSDTLKADNNTNTMYTLNCSNTHIKQACPKVSEPIPIYCA 221

DB 162 TVLKDKEKKQALFYVSDTLKADNNTNTMYTLNCSNTHIKQACPKVSEPIPIYCA 221

QY 181 TVLKDKEKKQALFYVSDTLKADNNTNTMYTLNCSNTHIKQACPKVSEPIPIYCA 239

DB 181 TVLKDKEKKQALFYVSDTLKADNNTNTMYTLNCSNTHIKQACPKVSEPIPIYCA 239

QY 222 PAGYAIKCNASAEFNGTKCNISVWTCRHGIRKPTVSTQILNGLTSKEKIRIMGNISD 281

DB 222 PAGYAIKCNASAEFNGTKCNISVWTCRHGIRKPTVSTQILNGLTSKEKIRIMGNISD 281

QY 240 PAGYAIKCNASAEFNGTKCNISVWTCRHGIRKPTVSTQILNGLTSKEKIRIMGNISD 299

DB 240 PAGYAIKCNASAEFNGTKCNISVWTCRHGIRKPTVSTQILNGLTSKEKIRIMGNISD 299

QY 282 SGKNIIVTLSSDIEITCVRPGNNTVQEMKIGPMANYSMALGTGSR--SRVAYCOYNTT 339

DB 282 SGKNIIVTLSSDIEITCVRPGNNTVQEMKIGPMANYSMALGTGSR--SRVAYCOYNTT 339

QY 300 SGKNIIVTLANTINITCERWT--NQSVEIPIGPMANYSMSVEKDKNTGSRSDCQNTS 358

DB 300 SGKNIIVTLANTINITCERWT--NQSVEIPIGPMANYSMSVEKDKNTGSRSDCQNTS 358

QY 340 EWEKALKNTAERYLELINT--EGNTTMIENRSDGSDVEVTHLHFNCHGEFFYCNTSEM 397

DB 340 EWEKALKNTAERYLELINT--EGNTTMIENRSDGSDVEVTHLHFNCHGEFFYCNTSEM 397

QY 359 EWTALQOEAERYLELNMNTGNTDNTVFNHS--TGDPPEVSLHFNCHGEFFYCNTSGM 417

DB 359 EWTALQOEAERYLELNMNTGNTDNTVFNHS--TGDPPEVSLHFNCHGEFFYCNTSGM 417

QY 398 FNYTFLCNGTNCNNTSINSANG---MPCKLQVYRSMRGGSLYAPPIGPNLTCLISH 454

DB 398 FNYTFLCNGTNCNNTSINSANG---MPCKLQVYRSMRGGSLYAPPIGPNLTCLISH 454

QY 418 FNYTFLCNGTNCNNTSINSANG---MPCKLQVYRSMRGGSLYAPPIGPNLTCLISH 477

DB 418 FNYTFLCNGTNCNNTSINSANG---MPCKLQVYRSMRGGSLYAPPIGPNLTCLISH 477

QY 455 ITGMILQMDAPWNTKTE-----ENRTSSSEPMKTCENIT 510

DB 455 ITGMILQMDAPWNTKTE-----ENRTSSSEPMKTCENIT 510

QY 478 ITGMILQMDAPWNTKTE-----ENRTSSSEPMKTCENIT 537

DB 478 ITGMILQMDAPWNTKTE-----ENRTSSSEPMKTCENIT 537

QY 511 IGTGTHREKRAVGLGMLFLGVLSAAGSTMGAATAATYVQHSVIGVQOQDNLRAIOA 570

DB 511 IGTGTHREKRAVGLGMLFLGVLSAAGSTMGAATAATYVQHSVIGVQOQDNLRAIOA 570

QY 538 IGTGTHREKRAVGLGMLFLGVLSAAGSTMGAATAATYVQHSVIGVQOQDNLRAIOA 597

DB 538 IGTGTHREKRAVGLGMLFLGVLSAAGSTMGAATAATYVQHSVIGVQOQDNLRAIOA 597

QY 571 QOELLRLSVGIRQLRARLLALETLIQNQQLNLNKGCKRLICHTSVKNWETWNTNIN 630

DB 571 QOELLRLSVGIRQLRARLLALETLIQNQQLNLNKGCKRLICHTSVKNWETWNTNIN 630

QY 598 QOELLRLSVGIRQLRARLLALETLIQNQQLNLNKGCKRLICHTSVKNWETWNTNIN 657

DB 598 QOELLRLSVGIRQLRARLLALETLIQNQQLNLNKGCKRLICHTSVKNWETWNTNIN 657

QY 631 QIWNGLTQWEDQIDNVSSTIYEEIOKAOVQOQNEKKLELDEWASLWNNLDTIKWLT 690

DB 631 QIWNGLTQWEDQIDNVSSTIYEEIOKAOVQOQNEKKLELDEWASLWNNLDTIKWLT 690


```

Db 1 MIVTKAMERKRNKLLWTLAYALITPCLSLQIATYAGVPVWEDATPVLFASDANL 60
QY 61 TSTEKNHNASQACVPTDPTPEHYLLNTVDFNENWENYVWQEQEDIIISLWQSLKPCI 120
Db 61 TSTEKNHNASQACVPTDPTPEHYPLHNTVDFNENWENYVWQEQEDIIISLWQSLKPCI 120
QY 121 QMTFMCIONNCTDIKNNTSGTENRTSSSENPMKTCFENITTVLKDDKKEKQALFYVDL 180
Db 121 QMTFMCIONNCTDIKNNTSGTENRTSSSENPMKTCFENITTVLKDDKKEKQALFYVDL 180
QY 181 TKLADNNTN---TWYTLNCSNTTIKQACPKVSEPIPIYCAPAGYAIKCNSEFNGTG 170
Db 171 MELNETSSNTKNSKMTNCSNTTIKQACPKVSEPIPIYCAPAGYAIKCNSEFNGTG 236
QY 237 GTGKCSNISVCTTGKIPVSTQILNGLTSLKSKIRIMGNISDSGKNIIVTLSSDIEI 296
Db 231 GTGTCRNITVCTTGKIPVSTQILNGLTSLKSKIRIMGNISDSGKNIIVTLSSDIEI 296
QY 297 TCVRGNNTVOEMKIGPMWYSNAL-GTGSNRVRVAYQVNTTWEKALKNTAERYLEL 355
Db 291 TCERP-QIDQEMKIGPMWYSNAL-GTGSNRVRVAYQVNTTWEKALKNTAERYLEL 348
QY 356 INNTGNTTMIENRSQSDGVETHLHFNCHGEFFYCNTESEMYTFLCNGTNCNTQSI 415
Db 349 VNNT-GSINMTFHS-SGGDLVTHLHFNCHGEFFYCNTESEMYTFLCNGTNCNTQSI 415
QY 416 NSANGMPCKLKQVYRSMRGSGGLYAPPVGNLTCTISHITGMILQMDAPWNKTEN--TF 473
Db 407 QGNNGTLPCKLRQVYRSMRGSGGLYAPPVGNLTCTISHITGMILQMDAPWNKTEN--TF 466
QY 474 RPIGDMKDIDWNLKFKYVYRVPKFSVAPTRPVIGTTHREKRAVGLMFLGVLS 533
Db 467 RPIGDMKDIDWNLKFKYVYRVPKFSVAPTRPVIGTTHREKRAVGLMFLGVLS 526
QY 534 AAGSTGAATALTVPQHSVIGKIVQOQDNLRAIQAOQELLRLSVGIRQLRALLALE 593
Db 527 AAGSTGAATALTVPQHSVIGKIVQOQDNLRAIQAOQELLRLSVGIRQLRALLALE 586
QY 594 TLQNOQLLWCKGRLCYTSVKKNETWRNTNTNQTNWGLTWQEDQIDNYSSTIY 653
Db 587 TLQNOQLLWCKGRLCYTSVKKNETWRNTNTNQTNWGLTWQEDQIDNYSSTIY 643
QY 654 EETQKAOQEOQNEKKLLEDEWASLWNLWDITKWLWYIKIITVIGALIGRIVMVLN 713
Db 644 EETQKAOQEOQNEKKLLEDEWASLWNLWDITKWLWYIKIITVIGALIGRIVMVLN 703
QY 714 LVNIRGQOPLSLQIPTRQOEAETPGTGEGGDEGRPLIPSPQGFLLPLTDLRTI 773
Db 704 LVNIRGQOPLSLQIPTRQOEAETPGTGEGGDEGRPLIPSPQGFLLPLTDLRTI 763
QY 774 ILWYHLLSNLSGRTQVISHRLGLWILGOKIIDACRICAAVHYWLOELKQSAATSLD 833
Db 764 ILWYHLLSNLSGRTQVISHRLGLWILGOKIIDACRICAAVHYWLOELKQSAATSLD 823
QY 834 TFVAVANWTDIILGIGRGLGILNIPRRVQFERSLL 873
Db 824 TFVAVANWTDIILGIGRGLGILNIPRRVQFERSLL 863

RESULT 8
ID O93091 PRELIMINARY; PRT; 860 AA.
AC O93091.
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
ENVELOPE GLYCOPROTEIN.
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-SAMPLE 193HA;
RA Hunt J.C., Golden A.M., Lund J.K., Gurtler L.G., Zekeng L., Obiang J.,
RA Kapteue L., Hampl H., Vallari A., Devare S.G.;
RT "Envelope Sequence Variability and Serologic Characterization of HIV-1
RT Group O Isolates from Equatorial Guinea."
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U82990; AAB94313.1;
DR INTERPRO; IPR000328;
DR INTERPRO; IPR000777;
DR PFAM; PF00516; GP120; 1.
DR PFAM; PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 860 AA; 96914 MW; A24B989AAE46DD8D CRC64;

Query Match 77.48; Score 3577.5; DB 12; Length 860;
Best Local Similarity 76.68; Pred. No. 7.3e-288;
Matches 671; Conservative 88; Mismatches 98; Indels 19; Gaps 7;

QY 1 MIVTKAMERKRNKLLWTLAYALITPCLSSQIATYAGVPVWEDATPVLFASDANL 60
Db 1 MIVTKAMERKRNKLLWTLAYALITPCLSSQIATYAGVPVWEDATPVLFASDANL 60
QY 61 TSTEKNHNASQACVPTDPTPEHYLLNTVDFNENWENYVWQEQEDIIISLWQSLKPCI 120
Db 61 TSTEKNHNASQACVPTDPTPEHYPLHNTVDFNENWENYVWQEQEDIIISLWQSLKPCI 120
QY 121 QMTFMCIONNCTDIKNNTSGTENRTSSSENPMKTCFENITTVLKDDKKEKQALFYVDL 180
Db 121 QMTFMCIONNCTDIKNNTSGTENRTSSSENPMKTCFENITTVLKDDKKEKQALFYVDL 180
QY 181 TKLADNNTN---TWYTLNCSNTTIKQACPKVSEPIPIYCAPAGYAIKCNSEFNGTG 229
Db 174 IKLEN-----STWYTLNCSNTTIKQACPKVSEPIPIYCAPAGYAIKCNSEFNGTG 229
QY 241 CSNISVVTCTGKIPVSTQILNGLTSLKSKIRIMGNISDSGKNIIVTLSSDIEITCVR 300
Db 230 CNNTVTCTGKIPVSTQILNGLTSLKSKIRIMGNISDSGKNIIVTLSSDIEITCVR 289
QY 301 PGNNQVQEMKIGPMWYSNALGTGSNRVRVAYQVNTTWEKALKNTAERYLEINTE 360
Db 290 PAIN--VQOVGLPLAVSYSLGVENITSRTAYCEYNTAAWERTLOQTAERWELVNTK 347
QY 361 GNTMTMFNSODGSDVETHLHFNCHGEFFYCNTESEMYTFLCNGTNCN-NTOSINSAN 419
Db 348 INVSIIFNOSGGGDAETHLHFNCHGEFFYCNTESEMYTFLCNGTNCN-NTOSINSAN 407
QY 420 GMIPCKLQVRSWVRGSGGLYAPPVGNLTCTISHITGMILQMDAPWNKTEN--TFRPIG 477
Db 408 TQIPCKLQVRSWVRGSGGLYAPPVGNLTCTISHITGMILQMDAPWNKTEN--TFRPIG 467
QY 478 GDMKIDWNLKFKYVYRVPKFSVAPTRPVIGTTHREKRAVGLMFLGVLSAAGS 537
Db 468 GDMKIDWNLKFKYVYRVPKFSVAPTRPVIGTTHREKRAVGLMFLGVLSAAGS 527
QY 538 TNGAATALTVPQHSVIGKIVQOQDNLRAIQAOQELLRLSVGIRQLRALLALELIQ 597
Db 528 TNGAATALTVPQHSVIGKIVQOQDNLRAIQAOQELLRLSVGIRQLRALLALELIQ 587
QY 598 NQQLNLWCKGRLCYTSVKKNETWRNTNTNQTNWGLTWQEDQIDNYSSTIYEIEI 657
Db 588 NQQLNLWCKGRLCYTSVKKNETWRNTNTNQTNWGLTWQEDQIDNYSSTIYEIEI 644
QY 658 KAQVQEQNEKKLLEDEWASLWNLWDITKWLWYIKIITVIGALIGRIVMVLN 717
Db 645 KAQVQEQNEKKLLEDEWASLWNLWDITKWLWYIKIITVIGALIGRIVMVLN 704
QY 718 IROGYQPLSLQIPTRQOEAETPGTGEGGDEGRPLIPSPQGFLLPLTDLRTI 777
Db 705 IROGYQPLSLQIPTRQOEAETPGTGEGGDEGRPLIPSPQGFLLPLTDLRTI 764
QY 778 YHLLSNLSGRTQVISHRLGLWILGOKIIDACRICAAVHYWLOELKQSAATSLD 837
Db 778 YHLLSNLSGRTQVISHRLGLWILGOKIIDACRICAAVHYWLOELKQSAATSLD 837

```

D	b	765	YHLLSNASGIQVTRHLGLGILWGKIIISACRVCVAANTQTWIQELQNSATSLDITAV	822
Q	y	838	AVANWTDDIIGLQIRGRGINIPRRVRQGFERSLL	873
D	b	825	AVANWTGGIISGIIQRIGRWNIIPRIQRLERILL	860
R	E	S	RESULT 9	
I	D	Q76576	PRELIMINARY; PRT; 877 AA.	
A	C	O76576;		
D	T	01-NOV-1996	(T=EMBLrel. 01, Created)	
D	T	01-NOV-1996	(T=EMBLrel. 01, Last sequence update)	
D	T	01-MAY-2000	(T=EMBLrel. 13, Last annotation update)	
D	E	ENVELOPE GENE (FRAGMENT).		
G	N	ENV.		
O	X	Human immunodeficiency virus type 1.		
O	C	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
O	X	NCBI_TaxID=11676;		
R	N	[1]		
R	C	SEQUENCE FROM N.A.		
R	C	STRAIN=VAU;		
R	X	MEDLINE=95065659; PubMed=7975221;		
R	A	Charneau P., Borman A.M., Quillent C., Guetard D., Chamaret S.,		
R	A	Cohen J., Remy G., Montagnier L., Clavel F.;		
R	T	"Isolation and envelope sequence of a highly divergent HIV-1 isolate:		
R	T	definition of a new HIV-1 group.";		
R	L	Virology 205:247-253(1994).		
D	R	ENBL: X80020: CAA56323.1; -		
D	R	INTERPRO: IPR000328; -		
D	R	INTERPRO: IPR000777; -		
D	R	PFAM: PF00516; GP120; 1.		
D	R	PFAM: PF00517; GP41; 1.		
K	W	Envelope protein.		
F	T	NON_TER		
F	T	NON_TERM		
F	T	SEQUENCE 877 AA; 99039 MW; 40C19E10D4981322 CRC64;		
Q	y	Query Match	{ 73.9%; Score 3414; DB 12; Length 877;	
Q	y	Best Local Similarity { 73.0%; Pred. No. 2.7e-274;		
Q	y	Matches 646; Conservative 85; Mismatches 134; Indels 20; Gaps		
Q	y	1 MIVTMRAAGKRNRKLGLYIVMALIIPCILSSQLYATVYAGVPVWEADAAPVLFCASDANL	60	
D	b	1 MTALMKAAGKRNRKLGIWCLIALIIPCILSCNLQLYATVTSVGVYEWEDAKPTLFCASDANL	60	
Q	y	61 TSTEKHNVASQACPVTDPDPHEYLNTVDNINVENTMWEOQEDIIISLDQSCLKPCI	120	
D	b	61 TSEQHNIAWTOACPVTDPSPNEYELKNVTKGNINWKNTIVDMQHEDIIDLWDQSCLKPCV	120	
Q	y	121 QMTFMCQMCTDKN--NNTSGTENRTSSSENPMTCFENITTVLKDKKKQAIFYYS	178	
D	b	121 QMTFLCVMQCTDKINSINTNGPLNSNKYE--VKQCDFNVTTVLKDAQEKNAIFYVI	178	
Q	y	179 DLTLADLNNTTN-TMYPLINCNSTTIKQCPKVSFEPIPIYYCAPAGYAIFRCNSAEFG	237	
D	b	179 DLVKI--NAISNETMYRLINCNSTTIQACPCKVSFEPIPIHYCAPAGCAIFRCNETGFG	236	
Q	y	238 TGKCSNISVTCGHGIKPTVSTOLINGTLUSKEKIRIMGNISDSOKNIIVTLSSDIET	297	
D	b	237 TGLCKNVTVTCTHGIRFTVSTOLIINGTLUSKGNITIMGKNISDSGENLIITNTITTA	296	
Q	y	298 CVRGNNOTVOEMKIGPMWYNALGTGSNRSVAVCOYNTTWEEKALKNTAERYLELIN	357	
D	b	297 CERPG-NQTIQIMAGPMWYSMALSNTKGDTAAACYNSATDNWALKNITERYLELVE	355	
Q	y	358 NTEGNTMIFNRSODGSDVEVTHLFHCHEFFCYCNTEMFNTFLC-----NGTCNMN	411	
D	b	356 YNQTDVTMKFG-NHSGEDAEVTFNFCHGEFFCYCNTRLNFHTFSCKKNWNKNINCTN	414	
Q	y	412 TQSIANSANGMIPCKLRKQVRSWMRGGSLYAPPIGNLTICISHITGMILQMDAPWNKT	-- 469	

Db 415 ISNSNCTQAIPECLRQVVDWVRGSSGLYAPPIPGNLVCRSNITGMILQLDTPWNKTHP 47
Qy 470 -ENTFRIGDDIKWRNELFKYKVVRKPFVSAPTPIARPVTGTGTHREKRAVLGMLF 528
Db 475 NSTTLRFGGGDKIWTQTLLKFVKVVRKPFVSAPTPTIARTTIGTRSHREKRAAGLAMLF 534
Qy 529 LGVLSAAGSTWGAAATALTATVOTHSVIKGIVQQDNLLRAIQAOQEILLRLSVGIOLRAR 588
Db 535 LGISLAAGSTWGAAATALTATVOTHLKYGVQQDNLLRAIQAOQHLLRPVSWGIRQLRAR 594
Qy 589 LLALETLIQNOLLNLNGCKGRLLICTYSVKWNTEWRNTTNINGNLTWQEWDDQIDNV 648
Db 595 LLALETFIQNOLLNLNGCANKRLLICTYSVAWKNTWGDGN--ESIWDELFWOHWDDQINNV 652
Qy 649 SSTIEEIOAKVOOEQRNEKKLLEDEWASLWNLWDITRWLTWYIKIAIIIVGALIGRVIV 708
Db 653 SSFYEIKIQAEOEQEKNEKEELLEDEWASIMNWLDITKWLTWYIKIAIIIVGALIGRVVV 712
Qy 709 MIVLNLYNRNKGYQPISLOIPPROQSEATPORTEGGGDGDRPRLLISPQGFLPLLYT 768
Db 713 MIVLNLYKNIRQGYQPISLOIPTQOAEEVTPSGTGEGGGEDRRRTWPULOQGFLLHLLYT 772
Qy 769 DLRTTIILWSYHLSNLISGTFQTVISHLRGLWLTLGOKIIDACRICAAVTHYWLMQELQKSA 828
Db 773 DLRTTIILWIYHLSNLASETKUHRHGUGLGMIIGORTIEARCLFAIKAIQYLWQLQELTS 832
Qy 829 TSLSDTFAVANWTDIIILGILORGLRNIPRRVROGFERSLL 873
Db 833 TNLLDTAVANWTDSTILIGISIGRGILNPRIHQGLERLLL 877

RESULT 10

Q79670 PRELIMINARY; PRT; 876 AA.

AC Q79670;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)

DE ENVELOPE PROTEIN GP120/GP41.
ENV.

OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
[1]
SEQUENCE FROM N.A.
RP MEDLINE-94149848; PubMed-8107219;
RA Gurtler L.G., Hauser P.H., Eberle J., Von Brunn A., Knapp S.,
RA Zekeng L., Tsague J.M., Kaptue L.;
RT "A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon."
JT R. Virol. 68:1581-1585(1994).
RL ENBL; L20571; AAA44864.1; -
DR INTERPRO; IPR000328; -
DR INTERPRO; IPR000777; -
DR PRAM; PF00516; GP120; 1.
FFAM; PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 876 AA; 99245 MW; A92F868E37522BAB CRC64;

Query Match : 73.2%; Score 3381.5; DB 12; Length 876;
Best Local Similarity 72.7%; Pred.No.1.3e+271;
Matches 649; Conservative 93; Mismatches 131; Indels 17; Gaps

Qy 1 MIVTMRMANGRENKRKLGLTYIWMALIIPCLS-SSQLATYTYAGVPVVEDAAPVLFCASDAN 59
Db 1 MTVMTRYMKYNRKRSWLSYAMALLIPCLSYSKQLATYTSYGPVWVEEAAPVLFCASDAN 60
Qy 60 LTSTEKNVNWASQCVPDTDPPHEHYLLTNVDFNFNIWENMYWQMOMEDIISLMDQSILKPC 119
Db 61 LTSTEQHNWASQCVPDPNPHEFPFGUNFTVDFNDINKNTVDMQMHEDIISLWEQSLKPC 120
Qy 120 IQMTFMCIQNMCTDIKNNNTSGETNRTSSENPMKTCEFNITVLXDKKKEKOALFYSD 179

```

Db 121 EKMFVLCVQMCVDLQTKNT--GLNETI---NEMRNCSEFNTVLTDKKQKQALFYSD 176
Qy 180 LTKLADNNTN-TWYTLNCHNSTTIKQACPKVSEPIPIYICAPAGYAFKCNKSAEFGT 238
Db 177 LSKVNSNAVNGTTVMTLNCSTTIKQACPKVSEPIPIHYCAPTGAFKCNKNDTENG 236
Qy 239 GKCSNISVVTCTHGIKPVSTQILNGLTSKRIKINGKINSDSGKNIIVTLSSDIETC 298
Db 237 GLCHNISVVTCTHGIKPVSTQILNGLTSKRIKINGKNITESAKNIIVTLNTPINNTC 296
Qy 299 VRPGNNOTVOEMKIGPMWYMSALGTGN---RSVAVCOYNTTEWEKALNTAERTLEL 355
Db 297 IREGTAE-VQDIYTGPMWRSTLKRSTNTSPRSVAICTYKNTWENALQOATIRYINL 355
Qy 356 INNTGNTMIFNRSQDSGVETHLPHNCHGEFFYCYNTSEMFTYFL-CNGTNCNNTQS 414
Db 356 VNOTE-NVTIIFSR-TSGDAEVSHLHFNCHGEFFYCYNTSGMFTYFNTCTKSGCQETK 413
Qy 415 IN-SANGMIPCKLKQVYRSMRGSGLYAPPICGNLTCISHTGMILQMDAPWKT-EN 471
Db 414 SNETKNGTIPCKLRQVYRSMRGSGLYAPPICGNLTCISHTGMILQMDAPWKT-EN 471
Qy 472 TRPTGGDMKDWRNELEFKYVVRKPSVAPTPIARPVIGTGT-HREKRAVGLGMLFLG 530
Db 474 TLRPVGGDMKDWRNELEFKYVVRKPSVAPTPIARPVIGTGT-HREKRAVGLGMLFLG 533
Qy 531 VLSAAGTMGAATALTAVTHSVIRKIGVVOQDNLLRAOAOQELLRLSVGIRQLRALL 590
Db 534 VLSAAGTMGAATALTAVTHSVIRKIGVVOQDNLLRAOAOQELLRLSVGIRQLRALL 593
Qy 591 ALETIONOOLLNAGCGRLCYTSVKNWETWNTTINQINWGLTWQDQOQIDNVSS 650
Db 594 ALETIONOOLLNAGCGRLCYTSVKNWETWNTTINQINWGLTWQDQOQIDNVSS 653
Qy 651 TYIEETQAOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 710
Db 654 IYDEIQAOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 713
Qy 711 VLVNVRNIRGQOPLSLQIPTRQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 770
Db 714 ILNVKNIRGQOPLSLQIPTRQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 773
Qy 771 RTIILWSHLNLSIGTQVISHRLGLWILGQIIDACRICAAVHYWQLQELQKSATS 830
Db 774 RTIILWSHLNLSIGTQVISHRLGLWILGQIIDACRICAAVHYWQLQELQKSATS 833
Qy 831 LIDTFAVANWTDIILGIRGLRGLINIPRRVQGFERSLL 873
Db 834 LIDTFAVANWTDIILGIRGLRGLINIPRRVQGFERSLL 876

RESULT 11
Q9QNX7 PRELIMINARY; PRT; 840 AA.
AC Q9QNX7;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMELrel. 14, Last annotation update)
DE ENV PROTEIN.
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=95ML84;
RA Montavon C., Bibollet-Ruche F., Robertson D., Kounare B., Mulanga C.,
RA Esu-Williams E., Toure C., Mboup S., Saman E., Delaporte E.,
RA Peeters M.;
RT "The identification of a complex A/G/I/J recombinant HIV-1 virus in
RT different west African countries.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; AJ245481; CAB53049.1; --
DR INTERPRO; IPR000328; --
DR INTERPRO; IPR000777; --
DR PFAM; PF00516; G120; 1.
DR PFAM; PF00517; G41; 1.
SQ SEQUENCE 840 AA; 94876 MW; 509ADB44F41E3B4 CRC64;

Query Match 51.6%; Score 2383; DB 12; Length 840;
Best Local Similarity 53.1%; Pred. No. 7.6e-189; Indels 48; Gaps 15;
Matches 458; Conservative 143; Mismatches 213;

Qy 20 IVALLIPICLSSQATVATVAGVPVWEDAAPVLFACASDANLSTEXHKNVWASQACVPTDP 79
Db 19 LIIGLAIICATSNLWTVYVGVVWEDATILFCASDAKAYSAEKHNWATHACVPTDP 78
Qy 80 TPHEYLLTNVDFNFWNVEQOEDIIISLDQSLKPCIQMTFCIQMNCNCTDIANNNT 139
Db 79 NPOEIPLENTVENFMKNMVMQOEHDIISLWDESLKPCVLTPLCVTLNCTNVTDHGI 138
Qy 140 SGTENTSSSENPMKTCFENITTVLADKKEKQALFYVSDTLKADNNTNTMYTLINCN 199
Db 139 N--NNNTVEGKEIKCSFNVTTEIKDKKKKERALFYTLVDVPINDG--NNSTYRLINCN 195
Qy 200 STTIKQACPKVSEPIPIYICAPAGYAFKCNKSAEFGTGCNSIVVCTHGIKPTVST 259
Db 196 TSIKQACPKVSEPIPIHYCAPAGYAFKCNKSAEFGTGCNSIVVCTHGIKPTVST 255
Qy 260 QLLINGLTSKERTIMGKNIIVTLSSDIITCVRPNNOTVQEMKIGP--MANY 318
Db 256 QLLINGLSAEDEIIRSENHTNNAKIIIVOLNKTVOIRCTRPSNN--TRKSIPLGQOAFY 314
Qy 319 SMALGTGNSRSVAYCOYNTTEWEKALNTAERYLELNNTEGNTTMTFNRSDGSDVEV 378
Db 315 --ATGDIIGDIRAHQNVSTANKEYLQNTTEKLQLL-----NTNITNPSA--GGDLEI 366
Qy 379 THLFNCHGEFFYCYNTSEMFTYFLNCHNNTQNSANGMIPCKLQVYRSMRGGS 438
Db 367 TTHSFNCHGEFFYCYNTSOLFNSI---PESNETDIIT-----LPCKIKOIVEMWQVQ 417
Qy 439 GLVAPPIPGNLTCISHTGMILQMDAPWKTENTFERPIGGMDKIWRNELFKYVVRVKP 498
Db 418 AYVAPPAGNITCISNITGLLTLTDTGNTNTSEIIRPGGMDRDNWRNELYKVKVVKIKP 477
Qy 499 FSVAPTPIARPVIGTGTTHREKRAVGLGMLFLGVLSAAGSTMGAATALTAVTHSVIRKIG 558
Db 478 LGIAPTEARRRVG---REKRAVGIGAPFLGFLGTAGSTMGAASTLTAVQVRLLSGIV 533
Qy 559 QOQNLRLRAOAOQELLRLSVGIRQLRGLRGLINIPRRVQGFERSLL 873
Db 534 QOQNLRLRAOAOQELLRLSVGIRQLRGLRGLINIPRRVQGFERSLL 876
Qy 619 WNETWRNTNINQINWGLTWQDQOQIDNVSSITYIEEIOKAOVQOQOQOQOQOQOQOQO 678
Db 594 WNASWSNRT--FNEIWNNTNIEWDREINNTTQOYISLIEESQOQOQOQOQOQOQOQO 652
Qy 679 LWNLDITKWLWIKIAIIVGALIGVRIYVIMLVNVRNIRGQOPLSLQIPTRQOQOQO 738
Db 653 LWNWEDISNLWIKIFIMVGLIGLIRVAVLSVNRVVRGQYSPSLQTLTLPNPAEVD 712
Qy 739 TPGRTGGGGGDEGRPLRIPSGOGLPLLYTLTDTITILWSYHLNLSN--LISTQIVSHRL 797
Db 713 RPGGIEEGGEGQGRNRSIRLVNGFLALAWDDLRLSLCFYHRLDRFVLIAARTVETLGR 772
Qy 798 G-----LWILGJKIIDACRICAAVHYWQLQELQKSATSIIIDTFAVANWTDIILGIQ 851
Db 773 GWELKYLWNL-----VC-----YWGQELKNSAISLIDTFAVANWTDVRIEVQ 818
Qy 852 RLGRGIILNIPRRVQGFERSLL 873
Db 819 RAFAVLNIPRRVQGFERSLL 873

```

```
RESULT 12
O40360 PRELIMINARY; PRT; 863 AA.
AC O40360;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRCPN;
RX MEDLINE=98216834; PubMed=9557756;
RA Blanchard A., Ferris S., Chamarret S., Guetard D., Montagnier L.;
RT "Molecular evidence for nosocomial transmission of human
RT immunodeficiency virus from a surgeon to one of his patients.";
RL J. Virol. 72:4537-4540(1998).
DR EMBL; U85914; AAC59351.1; -
DR INTERPRO; IPR000328; -
DR INTERPRO; IPR000777; -
DR PFAM; PF00516; GP120; 1.
DR PFAM; PF00517; GP41; 1.
KW Envelope protein.
FT MON_TER 1 1
SQ SEQUENCE 863 AA; 97628 MW; 4F635052B7397BDB CRC64;

Query Match 51.3%; Score 2370; DB 12; Length 863;
Best Local Similarity 52.3%; Pred. No. 9.5e-186;
Matches 464; Conservative 140; Mismatches 238; Indels 46; Gaps 14;

QY 5 MRANKRNRKL---GILYVVALIIPCLSSQLYATVYAGVPVWEDAPVLFCSADANLT 61
DB 1 VRGMORNRKQGNIGLY--LGILICNAADNLWTVYGVVWMDAKTTLFCASEAKAY 58
QY 62 STEKHNVASQACVPDTPHEYLNTVDFNFIWNTWVQMDIISLWDQSLKPCIQ 121
DB 59 ETEVHNWATACVPTDPNPOEILVNTFNFNWMDNNDVQMDIISLWDQSLKPCVK 118
QY 122 MTFQICMNCNCTDIKNNTSGTENKTS-----SENPKTCFENITVLKDKREK 171
DB 119 LTPCLVTLNCNITKSAFANDINSTAPLTSYIIEARRELTCSPNVTIELRDKYKV 178
QY 172 QALFVSDLTKLADNNTN-----TMYTLINCNTTKQACPKVSFEPIPIYCAPAGYA 226
DB 179 YSIFVQLDVQDNNSSNSNSYFLINCNASTITQACPKVSFEPIPIYCAPAGYA 238
QY 227 IFKNSAEFNGKCSNIVSVCTHGKPTVSTOLILNGTLSEKIRTMGNISDSGNI 286
DB 239 ILKCKDTFNGTGPCSNVSTVQCHGKIPVSTQLLNGSLAEKIMRTNKRSDSPGNI 298
QY 287 IVTLSSDIEITCRVPGNNTQVEMKIGP--MAWYSMALGTGNSRVRVAYCOYNTTEWEKAL 345
DB 299 IVQLIKPKVINKTRP--NNTRKSVHLGPGQAFY--ATGDIIGNVRQAFCTVNRTEKNLT 355
QY 346 KNTABRYLELNNTEGNTMTFNRSQDSQDVVEVTHLHNCHEGFFYCNTESENFYTLGN 405
DB 356 FNVSQLAEYFNKTE-----YFKESSGGDVEITTHSFNCRGEFFYCNTEGLFNSSFIQN 410
QY 406 GYNCHNNTOSINSANGMIPKLVQVRSVMRGSGLYAPPINLNTCTISHITGMILQMDAP 465
DB 411 GTSNDTND---TSANTTIPCRILQIVRMQRVQGMIAPIPGVITCVSNITGLLTGDGE 467
QY 466 WNKTEFTPIGGDMKDWRNELFKYKVRVYKVPFSVAPTPIARPVICTGTREKRAYVLG 525
DB 468 KNGINETRPIGGNWRNWSRLRYKVVRIEPLGVAPTRARRVV---EREKRAYVLG 523
QY 526 MFLGLVLSAAGSTGGAATALTVOHSHVIGVQOQDNLRAIQAOQELLKSLVWGIRQL 585
DB 524 AVFLGLAAGSTGGAATITLVOARQLLSIGVQOQSNLLRAIEAQOQMLRLTVWGIKQL 583
```

```
QY 586 RARLALLETLIQNOQLNLWGCGRGLICTSVKWNTRNTTINQIWNLTQWEDQOI 645
DB 584 QARVLAVERYLKQQLLGIWGCGRGLICTTAVPNSSWSN--KSFNELWNTWLTQWDEKEI 642
QY 646 DNVSTIYIEIQKAQVOOEQNEKKLLEDEWASLWNLDTIKWLVIKIAIIVGALIGV 705
DB 643 SNYSETIYRLIEDSQNQEKNEQDLSLDQWGLWNFDTIRLWLVKIFITMIVGGILG 702
QY 706 RIVMVLNLVRNIRQGYQPLSLQIPTRQOSEAETPGRTGEGGDEGRPLIPSPQGFPL 765
DB 703 RIIFAVLSIVNRVQGYSPLSFTLIPSPRGDPREEIEERGEGQDGRSIRLVNGFLAL 762
QY 766 LYTDRLTIILWSYHLLSN--LISGTQTVISHRLGLWTLGKIIDACRICAACAAHYWLQEL 824
DB 763 AWEDLRNLCLFSYRLDLALLAARTLDRLR--GSM-----EILKYLGNLVQYNGQEL 814
QY 825 QKSATSLIDTFVAVANWTDIILGIGLQRLGILNIPRRVRQGFERSL 872
DB 815 RNSAISLLNTAIVAEGRDRIEILQIRAGRAILHIPRIRROGAERIL 862
RESULT 13
Q9QRW2 PRELIMINARY; PRT; 860 AA.
AC Q9QRW2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE ENV.
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE8131;
RA Laukkanen T., Salminen M.O.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF107771; AAD49789.1; -
DR INTERPRO; IPR000328; -
DR INTERPRO; IPR000777; -
DR PFAM; PF00516; GP120; 1.
DR PFAM; PF00517; GP41; 1.
SQ SEQUENCE 860 AA; 97757 MW; 2923610F7D7ACAC9 CRC64;
```

Query Match 51.0%; Score 2356.5; DB 12; Length 860;
Best Local Similarity 51.8%; Pred. No. 1.2e-186;
Matches 461; Conservative 146; Mismatches 232; Indels 51; Gaps 16;

```
QY 5 MRANKRNRKLGL---YIVMALIIPCLSSQLYATVYAGVPVWEDAPVLFCSADANLT 61
DB 1 MRVGTORNCOLLNMGIMILGMIICSTAENLWTVYGVVWMDAETTLFCASDAKAY 60
QY 62 STEKHNVASQACVPDTPHEYLNTVDFNFIWNTWVQMDIISLWDQSLKPCIQ 121
DB 61 EKEVHNWATACVPTDPNPOEILVNTFNFNWMDNNDVQMDIISLWDQSLKPCVQ 120
QY 122 MTFQICMNCNCTDIKNNTSGTE--NRTSSSNPMKTCFENITVLKDKKQKQALFYVSD 180
DB 121 LTPCLVTLNCNNTANTNSTANLTSVKGEMRNCNFTITELRDKKRVSYLFYKLDI 180
QY 181 TKLADN-----NTNTMYTLINCNTTITKQACPKVSFEPIPIYCAPAGYAFKC 230
DB 181 VKNNKNSFRGNSSGSSDRYLINCNTSAITQACPKVSFEPIPIYCAPAGYAFKLC 240
QY 231 NSAEFNGTGCNIVSVCTHGKIPVSTOLILNGTLSEKIRTMGNISDSGNIIVTL 290
DB 241 NEDEFNGTGCNIVSVCTHGKIPVSTOLILNGTLSEKIRTMGNISDSGNIIVTL 300
QY 291 SSDIEITCRVPGNNTQVEMKIGP--MAWYSMALGTGNSRVRVAYCOYNTTEWEKALNTA 349
```



```

Db 301 TKPVELICTRP--NNNTRKSIRIGPQAFYGMGDIIGDIRK--AHCNVSRSKWNELAKVA 357
QY 350 ---EYLELINNTGNTMIENRSDGSDVEVTHLHFNCHGEFFCYNTSEMFNYELCNG 406
Db 358 IOLRKY-----WNTTILFTNS--SGDLEITHTSFNCGGEFFCYNTSGLFNSWEND 408
QY 407 TNCN-NTOSINSANGMIPCKLKQVRSWMRGSGLYAPPPIPNLTCTISHITGMILQMD-A 464
Db 409 TKVNYNESNTIIT--LQCRKIQIINWQRTGQATYAPPPIGVQCRSNTINGLLTRDGG 466
QY 465 PNWKTEN--TFRPIGDMKDIWRNELFKYKVVVRPFSVAPTPPIARPVIGTTHREKRAVG 523
Db 467 VTNNNTNETFRPGDMDNRSELYKYVVKLEPLGVAPTAKARRV-----KREKRAVG 522
QY 524 LGMLFLGVLSAAGSTMGAATAAFTVQTHSVLKGIVQOODNLLRAIQAOQELLRSVWGIR 583
Db 523 LAAYFFGLGAGSTMGAASITITVQARQLLGGIVQOONLLRAIEAQOHLRLTVWGIR 582
QY 584 QLRLALALETLIQOOLLNLWGCKGLICYTSVKWNETWRTNTINQINGNLTWQEDQ 643
Db 583 QLQARVMAVERYLADQOLLGTCGSGKIICHTTAVPNSTWSN--KSYTQIWNMTWLQWDK 641
QY 644 QIDNVSSTIYEIEOKAQVQOQONKEKLELDEWASLNNWLDITKWLWIKIATIIIVGALI 703
Db 642 EISNYTDIIYQIIEESQYQEKNEKLELDEKLANWLNWFDISNWLMTIKIFIMVGLI 701
QY 704 GVRIVMVLNVRNIRGQYPLSLQIPTROQSEAEPTCGTGGGDRGRPLIPSGFL 763
Db 702 GLRIVFVLTVIKRVRGYSPLSQIHTPSPDRPRGRIEESGDRGRSIRLVSGFL 761
QY 764 PLYTDLTILWSYHLLSNISQITQVISHRLGLWMLGKIIDACRICAAVHYWLOE 823
Db 762 ALANDLRLSCFSYHRLRDISIATRIVELGGLGKLVGLNLL-----YWIRE 810
QY 824 LOKSATSILDTFAVANWNTDDIILGQRLGRLGILNIPRRVQGFERSLL 873
Db 811 LKISALSLEDTAIAVAGWTDVIEIGRIGRAILHIPRIRQGFERALL 860

RESULT 14
Q9WLJ8 PRELIMINARY; PRT; 862 AA.
AC Q9WLJ8;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN.
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N1144;
RA McCutchan F.;
RT Limited diversity of subtype E envelope sequences from recent,
RT seroincident cases of human immunodeficiency virus type 1 infection in
RT Thailand.
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF070703; AAC99347.1;
DR INTERPRO; IPR000328;
DR INTERPRO; IPR000777;
DR PFAM; PF00516; GP120; 1.
DR PFAM; PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 862 AA; 97667 MW; 256A6E2C8069FC15 CRC64;

Query Match 51.08; Score 2356; DB 12; Length 862;
Best Local Similarity 51.68; Pred. No. 1.4e-186;
Matches 451; Conservative 144; Mismatches 229; Indels 50; Gaps 15;
QY 20 IVMALIIPCLSSQLYATVYAGVPWEDAAPVLCASDANLTSTKHNWNASQACVPTDP 79

```

```

Db 19 LILGLVICSASNLWTVYVGVVWRDADTTLFCASDAKAKAHETEAHNIWATHACVPTDP 78
QY 80 TPHEYLITNTVDFNFWENWVQMOQEDIIISLDQSLKPCIQMTFMCIONCNDIKNNNT 139
Db 79 NPOEIHLENTFNFMKNWVQMOQEDVLSLDQSLKPCIKLTCLVTLNCDSDAKFKV 138
QY 140 SGTENRTSSSEN--PMKTCENITTVLKDKEKKKQALFYVSDLTKLADNNTNTMYTLI 196
Db 139 STINDSNIIGNITDEVRNCTFNMTTELKDKKONVRALFKLDIVQL--ENNNSREYRLI 197
QY 197 NCNSTTIKQACPKVSEPIPIYCAPAGYAFKCSAEFNGTGKCSNISVVTCTHGIKPT 256
Db 198 NCNSTVIKQACEKISFDPPIHYCTPAGYAILKCNKDNFNGTGPKNVSSVQCTHGIKPV 257
QY 257 VSTQILLNGTSLSEKTRIMGNKNSDSGKNIIVLSSDIEITCVRPNGNQVQEMKIGP-M 315
Db 258 VSTQILLNGSLAEETIIRSENLANNAKTIIVHLNKSVEINCTRPSNN--TRTGITIGPGQ 316
QY 316 AWYSMALGTGNSRVAYCOYNTTEWEKALKNTAERYLELINNTEGNTTIFNRSODGSD 375
Db 317 VFRTGDTGDIRK--AYCEVNGTKWEILKQVTEKLEHEFNT-----IVQPPSGGD 368
QY 376 VEYTHLHFNCHGEFFCYNTSEMFNYTL-----CNGTNCNNTQSINSANGMIPCKLKQ 428
Db 369 LEITMHFNCHGEFFCYNTSRLENFTWIGNETMERCNGTGCNGTI-----ILPCKIKQ 421
QY 429 VVSWRGGSLYAPPIPNLTCTISHITGMILQMDAPWKTEN--TERPIGDMKDIWRNE 487
Db 422 IINMGAGAGATAPPYIRGISVCSNITGILLTRDGGANNASNETFRPGGNIKDNWRSE 481
QY 488 LFYKVVVRPFSVAPTPIARPVIGTTHREKRAVGLMFLGVLSAAGSTMGAATAALT 547
Db 482 LYKVKVQIEPLGVAPTAKARRV-----QREKRAVGAMIFGLNAGSTMGAASITLT 537
QY 548 VQTHSVKIGVQOODNLLRAIQAOQELLRSVWGIRQLRLALALETLIQOOLLNLWGC 607
Db 538 VQARQLLGGIVQOOSNLLRAIEAQOHLQLTVYGIKQLQARVLAVERYLKNQKFLGLWGC 597
QY 608 KRLICYTSVKWNETWRTNTINQINGNLTWQEDQIDNVSTIYEIEOKAQVQEQNE 667
Db 598 SGKIICHTTAVPNSTWSN--RSFEEIWNNTWIEWEREISNYTSQIYEILTSESQOQDRNE 656
QY 668 KLELDEWASLNNWLDITKWLWIKIATIIIVGALIGVRIVMVLNVRNIRGQYPLSL 727
Db 657 KDLELDKASLANWFDITNWLWYIKIFIMVGLIGLRIIFAVLSIVNVRGYSPLSL 716
QY 728 QIPTROQSEAEPTCGTGGGDRGRPLIPSPQGFPLPYLTDLRTILWSYHLLSN-LIS 786
Db 717 QPTTHQREPDPRPERIEEGGEGGQDRSVRLVSGFLSLAWDDIRSLCLFLYHRLRDFILI 776
QY 787 GTQTV-----ISHRLGLWMLGKIIDACRICAAVHYWLOELQKSATSILDTFAVAV 839
Db 777 ATRVELLGHSSSLKGLRRG-W-----EGLKYLKLLIYWGQELKISALSLLDATAIAY 828
QY 840 ANWTDIILGILQSLGRLNIPRRVQGFERSLL 873
Db 829 AGWTDVIEAQVWRAILHIPRIRQGFERALL 862

RESULT 15
Q9WLJ1 PRELIMINARY; PRT; 859 AA.
AC Q9WLJ1;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN.
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]

```


[illegible]

